

GROUP DELAY TRACKING WITH THE SYDNEY UNIVERSITY STELLAR INTERFEROMETER

Thesis submitted for the degree of
Doctor of Philosophy

by

Peter Roderick Lawson



Chatterton Astronomy Department
School of Physics
University of Sydney
Australia

October 1993

Summary

In the Sydney University Stellar Interferometer (SUSI), fringes are detected when the optical path-difference between the arriving wavefronts is less than the coherence length for the observation. While observing a star, it is necessary to compensate not only for the changes in pathlength due to the earth's rotation, but also for those due to the effects of atmospheric turbulence. Apertures smaller than r_0 are used and active tilt-compensation is employed so that the wavefronts in the combined beams are made parallel. The remaining path-difference between the wavefronts must then be tracked for an accurate calibration of the fringe visibility. Improved tracking allows larger bandwidths to be used and therefore improves the sensitivity of the instrument.

In this thesis a fringe tracking system is developed for SUSI based on group delay tracking with a PAPA detector. The method uses short exposure images of fringes detected in the dispersed spectra of the combined starlight. The number of fringes across a fixed bandwidth is directly proportional to the path-difference between the arriving wavefronts, and a Fast Fourier Transform is used to calculate the spatial power spectrum of the fringes, thereby locating the delay. Several topics are developed in the thesis. The visibility loss due to a non-constant fringe spacing on the detector is investigated, and the improvements obtained from rebinning the photon data are shown. The low light level limitations of group delay tracking are determined, with emphasis on the probability of tracking error, rather than on signal-to-noise performance. Experimental results from both laboratory studies and stellar observations are presented. These show the first closed-loop operation of a fringe tracking system based on observations of group delay with a stellar interferometer.

A new photon counting PAPA detector is also described, which was developed for use in this work. The design principles of the PAPA camera are outlined, and the potential sources of image artifacts are identified. These artifacts arise from the use of optical encoding with Gray coded masks. The new camera is distinguished by its mask-plate, which was designed to overcome artifacts due to vignetting. New lens mounts are also presented which permit a simplified optical alignment without the need for tilt-plates. The performance of the camera is described and its images are shown to be free of the effects of vignetting.

Declaration of Originality

To the best of my knowledge, this thesis contains no copy or paraphrase of work due to any other person, except where duly acknowledged. None of the remaining work has been presented for any degree at the University of Sydney or elsewhere.

Peter R. Lawson.

Publications

Some of the material in this thesis is based on the following publications:

W.J. Tango, J. Davis, P. Lawson, A.J. Booth (1992) “Fringe detection and tracking in SUSI,” ESO Conference on *High-Resolution Imaging by Interferometry II*. (European Southern Observatory: Garching, Germany), 1267–1269.

Peter R. Lawson and Andrew J. Booth (1994) “The fringe tracking servo in SUSI,” IAU Symposium No. 158 on *Very High Angular Resolution Imaging*. J.G. Robertson and W.J. Tango, Eds., 184–186.

Peter R. Lawson (1994) “Artifacts in PAPA camera images,” *Appl. Opt.*, **33**, 1146-1153.

Peter R. Lawson (1994) “The Sydney University PAPA camera,” *Proc. Astron. Soc. Aust.*, **11**, No. 1, 50-54.

Acknowledgements

I would like to begin by thanking my parents and my brother. Their support and encouragement has always been of great comfort, even from the other side of the planet.

Thanks are due to Clive Standley and Pete Nisenson for introducing me to the PAPA camera and for the hospitality they showed me during my stay at Harvard. The Sydney University PAPA camera would not have been possible without them. My thoughts concerning its design and the understanding of camera artifacts were influenced by discussions with Wes Traub, Chris Dainty, Fred Reavill, Bill Robinson, and Jim Beletic. I am also grateful for the hospitality shown to me at Imperial College and Georgia State University when I was working through PAPA problems. The alignment of the optics was made easier with the help of Derek McKay, who was also responsible for laying the tracks of the new analog electronics. For the design of the electronics themselves, I am indebted to Hank Bennis for his expertise, his good engineering, and his patience; and also Tony Cañas, whose detailed notes I took to heart. I am grateful to Mark Colavita for sending me his camera software and for information about the PAPA used in the Mark III. With various camera parts I was fortunate to have the help of Mick Paterson, John Pasiut, Jim Sullivan, Jim Hayes, Ken Weigert, and particularly Fred Peterson, with whom I discussed many of the new designs.

I benefited from discussions with Bill Tango, Robert Minard, and Bob Shobbrook concerning various aspects of the project. Erik Thorvaldson and John Giovannis helped me during my observing runs at SUSI, and Erik in particular with waterworks for the PAPA. I would especially like to thank Andrew Booth and Steve Owens for the improvements they made to the optical path compensator. Thank you Steve for many memorable abseils and bushwalks, which didn't add to this work in the least. Theo ten Brummelaar, rescued me on innumerable occasions with his knowledge of both the CHIP interface and the Register Transfer Protocol, and, along with Michael Hrynevych, provided a most enjoyable work environment. I am grateful for their friendship, the road trips, and the good times we had.

Most importantly, I would like to thank my supervisor, John Davis, for his support throughout this project and for guidance when it was required. I take pleasure in acknowledging the support of a Sydney University Postgraduate Research Award and the Commonwealth Scholarship and Fellowship Plan. SUSI has been funded jointly by the Australian Research Council and the University of Sydney with additional support from the Pollock Memorial Fund and the Science Foundation for Physics within the School of Physics.

Contents

I	Introduction	1
1	Modern Michelson Stellar Interferometry	3
1.1	Stellar Interferometry	3
1.1.1	Van Cittert—Zernike theorem	4
1.1.2	Path compensation and fringe tracking	5
1.2	The Sydney University Stellar Interferometer	6
1.3	The Optical Pathlength Compensator	8
1.4	Motivation for this Work	9
1.5	Outline of the Thesis	10
2	Fringe Detection in Stellar Interferometry	11
2.1	Fringe Tracking with Wavefront Tilt Compensation	11
2.2	The Interference of Quasi-Monochromatic Light	12
2.3	The Coherence Envelope	13
2.4	Measurements of Delay Curves	15
2.5	Pathlength Modulation	16
2.6	Channeled Spectra and Group Delay Tracking	18
2.6.1	Model fitting by least-squares	19
2.6.2	Optimal and Kalman filtering	20
2.6.3	Spectrum analysis and the Fast Fourier Transform	21
2.7	Conclusion	23
II	The PAPA Camera	24
3	The Design of PAPA Cameras	26
3.1	Introduction	26
3.2	Address Decoding with Optical Masks	28
3.2.1	Principle of operation	28
3.2.2	Optical design	29
3.2.3	Alignment of Gray coded masks	29
3.3	Design Parameters	31
3.3.1	Image tubes and photon counting	33

3.3.2	Magnification tolerances	36
3.3.3	The field size of images	37
3.4	Historical Background	37
3.4.1	Papaliolios and Mertz (1982)	37
3.4.2	Papaliolios, Nisenson, and Ebstein (1985)	37
3.4.3	Adaptive Optics Associates (1988)	39
3.4.4	Standley & Nisenson (1989)	40
3.5	Artifacts in PAPA Camera Images	43
3.5.1	Alignment errors	43
3.5.2	Discriminator setting errors	47
3.5.3	Vignetting errors	49
3.5.4	The sum of all masks	50
4	The Sydney University PAPA Camera	53
4.1	Guidelines for the New Design	53
4.2	The Mask Plate	56
4.3	The Mechanical Design	56
4.3.1	The lens-mask assembly	56
4.3.2	Mounting of the field lenses	57
4.3.3	The realignment of the optics of the camera	57
4.4	The New Analog Electronics	62
4.5	Software and Computer Interface	62
4.6	Performance Tests	62
4.6.1	Flat field	62
4.6.2	Detective quantum efficiency	63
4.6.3	Resolution	66
4.6.4	Linearity and dark count	67
4.7	Summary	70
III	Theory of Group Delay Tracking	72
5	Delay Tracking with the FFT	74
5.1	The Channeled Spectrum	74
5.2	Periodogram Estimate of the Delay	75
5.3	Outline of Part III	78
6	The Detector Response	80
6.1	The Transfer Function of an Array Detector	80
6.1.1	Aliasing and the delay envelope of a pixel	81
6.1.2	The resolution in delay	83
6.1.3	Zero padding the array data	86
6.2	Visibility Loss Due to Non-Constant Dispersion	87

6.2.1	Linear mapping	87
6.2.2	Non-linear mapping	88
6.2.3	Grism spectrometer	91
6.2.4	Gridding, remapping, and interpolation	91
6.2.5	Power spectra with non uniform sample spacing	91
6.2.6	Rebinning of data	92
6.3	Conclusion	96
7	Fringe Detection at Low Light Levels	97
7.1	The Channeled Spectrum with Photon Noise	99
7.2	The DFT Power Spectrum Signal-to-Noise Ratio	100
7.3	The DFT Modulus Signal-to-Noise Ratio	107
7.4	The Probability of Error in Tracking	110
7.5	Extreme Low light level failure of the DFT approach	115
7.6	Conclusion	116
8	Delay Tracking and Atmospheric Turbulence	117
8.1	Visibility Loss Due to Fringe Motion	118
8.1.1	Coherent fringe detection	118
8.1.2	Incoherent fringe integration	119
8.2	Temporal Power Spectrum of Phase-Difference	120
8.2.1	Kolmogorov Turbulence	121
8.3	Previous observations of white-light fringe motion	124
8.4	Active and Passive Tracking of Fringe Motions	125
8.4.1	Active fringe tracking	125
8.4.2	Passive observations of delay	126
8.5	Conclusion	127
8.5.1	Active vs. passive interferometry	128
8.5.2	Group delay tracking and passive interferometry	128
IV	Experiments in Delay Tracking	130
9	Data Processing	132
9.1	Data Acquisition	132
9.2	The Calculation of the Periodogram	134
9.3	Integration of Power Spectra	137
9.4	Locating and Filtering the Delay	139
9.4.1	Locating the peak in the power spectrum	139
9.4.2	Tracking the delay	139
9.5	Passive Tracking	142
9.6	Software and Hardware	142
9.7	Summary	143

10 Performance Tests	144
10.1 Laboratory Measurements in Sydney	144
10.2 Tests using SUSI in Autocollimation	149
10.3 Conclusion	157
11 Stellar Observations	158
11.1 Observations of α CMa	158
11.2 Observations of α PsA and α Eri	166
11.3 The Path Compensating Servo	173
11.4 Conclusion	176
12 Conclusions	177
12.1 The PAPA Camera	178
12.2 Group Delay Tracking	180
Appendices	181
A Photographs	183
B Design of the Gray Coded Mask Plate	185
B.1 A Description of the Hexagonal/Square Array	185
B.2 Mask Descriptions	185
B.3 The Mask Plate	189
C Alignment of the PAPA Camera	190
C.1 Tilt Plates and Lens Rotation	190
C.1.1 Tilt-plates	190
C.1.2 Zoom-lens rotation	191
C.2 Apparatus and Preparation	192
C.2.1 Autocollimating the main lens	192
C.2.2 Lens-mask allocation	192
C.3 The Alignment Procedure	193
C.4 The Spectrum of the Image Tube Phosphor	194
D List of Parts for the PAPA Camera	195
D.1 Introduction	195
D.2 Camera Parts	196
D.2.1 1st Generation 25mm Varo Tube	196
D.2.2 25mm MCP 2nd Generation Varo Tube, Model 3603	196
D.2.3 Related Parts for Intensifiers	197
D.2.4 Camera Base	197
D.2.5 Lens-Mask Assembly	198
D.2.6 Camera Back End	198

D.3	List of Mechanical Drawings	199
D.4	Notes and Comments	200
D.4.1	The Image Intensifiers	200
D.4.2	The Collimating Lens	200
E	The Prism Spectrometer	202
E.1	Prism Equations	202
E.2	Calibration of the Spectrometer	203
F	Spectrum Analysis and the DFT	204
F.1	The Fourier Transform	204
F.2	Some Useful Functions and Fourier Transforms	205
F.2.1	Fourier transform relationships	205
F.2.2	The Rectangle function, $\Pi(t)$	206
F.2.3	The Sampling or Replicating function, $\text{III}(t)$	206
F.3	A Note on Sampling Theory	207
F.3.1	The reconstruction of a sampled signal	207
F.4	The Discrete Fourier Transform	208
F.5	Artifacts Produced by the DFT	208
F.5.1	Resolution and bandwidth	209
F.5.2	Aliasing	209
F.5.3	Spectral leakage	209
F.5.4	The picket-fence effect	210
G	Wave Propagation in Random Media	213
G.1	Introduction	213
G.2	Spatial Covariances of Log-Amplitude and Phase	214
G.3	Temporal Power Spectra of Log-Amplitude and Phase	216
G.3.1	Temporal power spectra in a homogeneous medium	216
G.3.2	Temporal power spectra in a smoothly varying medium	217
G.4	Temporal Power Spectra of Everything	218
G.5	Summary	218
	Bibliography	219

List of Figures

1.1	The Sydney University Stellar Interferometer	7
3.1	The layout of the PAPA camera optical components.	27
3.2	Gray coded masks for 4-bit address decoding.	30
3.3	A comparison of binary and Gray code.	31
3.4	The alignment chart for the Gray coded masks.	32
3.5	Previous lens arrays for PAPA cameras	38
3.6	A flat-field from the Standley & Nisenson camera.	41
3.7	An integrated cross section along the x axis.	42
3.8	An integrated cross section for the y axis.	42
3.9	A flat field from the Sydney University camera.	44
3.10	A cross section through the x axis.	45
3.11	A cross section through the y axis.	45
3.12	Flat field with defocus artifacts.	46
3.13	Flat-field with a discriminator error in the X1 channel.	48
3.14	Flat-field with a discriminator error in the X2 channel.	48
3.15	Flat-field with a vignetting error in one mask.	50
3.16	The sum of all masks.	52
3.17	An image with light level problems.	52
4.1	Possible lens arrays for the rebuilt camera.	54
4.2	The Gray coded mask plate of the new camera.	55
4.3	Design drawing of the array lens mount.	58
4.4	Design drawing of the support for the mask plate.	59
4.5	Design drawing of the photomultiplier tube housing.	60
4.6	Design drawing of the mounting plate for the lens-mask assembly.	61
4.7	The quantum efficiency of the Gen II photocathode.	63
4.8	N2 photomultiplier vs camera pixels.	64
4.9	An enlargement of the spectrum of α CMa.	64
4.10	The full spectrum of α CMa with the data of April 22.	65
4.11	An image of a resolution chart.	67
4.12	An image of finely spaced line pairs.	68
4.13	A graph of the linearity of the PAPA camera.	69
4.14	An image of the logo of the Sydney University Stellar Interferometer.	70

5.1	Some example fringes and their Fourier transform.	76
5.2	Spatial frequency response with dispersion constant in wavenumber. .	77
6.1	Aliasing and the delay envelope illustrated.	82
6.2	The DFT of a fringe for which the dispersion is constant.	84
6.3	The effect of averaging and zero padding.	85
6.4	Spatial frequency response with non-constant dispersion.	89
6.5	The peak height and location in the power spectrum with no rebinning.	90
6.6	The detector response after the data have been rebinned.	93
6.7	Spatial frequency spectrum after remapping for dispersion.	94
6.8	The peak height and location in the power spectrum after rebinning. .	95
7.1	Examples of photon rich and photon starved data.	98
7.2	Biased and unbiased estimates of squared-amplitude.	102
7.3	Integrated biased and unbiased estimates of squared-amplitude. . . .	103
7.4	SNR for integrated frames of unbiased power spectrum estimates. . . .	105
7.5	Contours for the probability of tracking loss.	112
7.6	The probability of tracking error for $ \gamma = 1.0$	113
7.7	The probability of tracking error for $ \gamma = 0.5$	114
8.1	Visibility loss due to fringe motion during coherent integration.	119
8.2	The power spectrum of phase-difference fluctuations.	122
9.1	The data flow through the Data Acquisition module.	133
9.2	The data format through the processors.	135
9.3	The data flow through the FFT processor and the recursive integration.	136
9.4	The frequency response of the single-pole recursive filter.	138
9.5	The frequency response of the moving average rectangular window. . .	141
10.1	The Michelson interferometer used for laboratory tests.	145
10.2	A sinusoidal delay pattern recorded in the lab.	147
10.3	The same as the previous pattern but with a filter.	147
10.4	Another trace of delay with a higher amplitude.	148
10.5	The previous trace with a moving rectangular filter.	148
10.6	Beam combining for autocollimation with SUSI.	150
10.7	The response of the tracking as a function of pathlength.	152
10.8	An envelope of delay measurements from tests at SUSI.	153
10.9	Four power spectra from autocollimation tests: 1-1 no pixel averaging.	154
10.10	Four power spectra from autocollimation tests: 2-1 pixel averaging. . .	155
10.11	The mean noise level as a function of the number of photons per frame.	156
11.1	Position errors of OPLC carriage at $-101 \mu\text{m/s}$ and $-200 \mu\text{m/s}$	159
11.2	The path-difference changes with α CMa on April 3, 1993.	161
11.3	Further traces of path-difference changes observed on April 3, 1993. .	162

11.4	Path-differences observed with α CMa on April 4, 1993.	163
11.5	Cross section of previous image showing the effect of averaging.	164
11.6	Power spectra recorded while observing α CMa.	165
11.7	Position errors of OPLC carriage in September 1993.	166
11.8	The group delay wander observed with α PsA	168
11.9	Another example of group delay wander observed with α PsA	169
11.10	Fringe motions observed with α Eri on Sept. 2 and 3, 1993.	170
11.11	Filtered fringe motions observed with α Eri on Sept. 4, 1993.	171
11.12	Filtered fringe motions observed with α Eri on Sept. 4, 1993.	172
11.13	Test of the fringe tracking servo with α Eri on Sept. 7, 1993.	174
11.14	Test of the fringe tracking servo with α Eri on Sept. 7, 1993.	175
12.1	Number of photons-per-frame for PAPA camera vs stellar B magnitude.	179
B.1	The layout of the Gray coded masks.	186
B.2	The X0 and Y0 Masks	187
B.3	The X1 and Y1 Masks	187
B.4	The X2 and Y2 Masks	188
B.5	X and Y Masks of order 3, 4, 5, 6, and 7	189
F.1	The picket-fence effect.	211

Part I

Introduction

Chapter 1

Modern Michelson Stellar Interferometry

The techniques of optical interferometry will profoundly change our understanding of astronomy. High angular resolution stellar interferometry provides the only means of determining many of the fundamental properties of stars (Davis, 1985a). This includes stellar radii, surface fluxes, effective temperatures, and thus luminosities; the modelling of stellar atmospheres, stellar rotation, limb darkening, binary star separations, and the direct calibration of the Cepheid period-luminosity relationship. The scientific progress in this field until the 1980's has been described in a comprehensive survey by McAlister (1985). He states that if long-baseline optical interferometers reach their expected sensitivities, they will trigger a revolution in how astronomers view the Universe.

This thesis describes the development of a fringe tracking system for the Sydney University Stellar Interferometer. As will be shown, fringe tracking is an essential component of modern long-baseline stellar interferometry. Also described here is the design of a new photon counting PAPA camera, which was developed and tested for use with the fringe tracking system. In this chapter the principles of stellar interferometry are reviewed, the motivation for the work is described, and the structure of the thesis is outlined.

1.1 Stellar Interferometry

The methods of astronomical interferometry have been reviewed by several authors (Justice et al., 1985; Thompson et al., 1986; Clark 1989). Although these reviews are primarily intended to describe radio interferometry, the principles that they discuss are equally valid in the optical regime. This is true of the van Cittert—Zernike theorem which relates the measured data to the structure of the source. The assumptions of

the theorem will be reviewed to illustrate the role of path compensation in a modern Michelson interferometer.

1.1.1 Van Cittert—Zernike theorem

A typical stellar interferometer uses a pair of collecting apertures that are located at fixed positions on the surface of the earth, denoted here by the vectors \mathbf{p}_1 and \mathbf{p}_2 . Behind the apertures are mirrors which redirected the light to a laboratory, travelling an additional distance s_1 from \mathbf{p}_1 , and s_2 from \mathbf{p}_2 , until it is combined to form interference fringes. Measurements are made of the contrast of the fringes, and perhaps also an estimate of the fringe phase. This is averaged over a time during which the fringes appear stable and is then recorded. The van Cittert–Zernike theorem relates the fringe measurements to the structure of the source.

The theorem treats only quasi-monochromatic waves, and makes several assumptions about the source and the propagating medium (Clark, 1989). It is assumed that the source is a long way away, has no depth, and lies on the celestial sphere—a sphere with an almost infinite radius which is centered on the observer. Within that sphere there are no other sources of electromagnetic radiation: the fields propagate through empty space from the surface of the sphere. The radiation is treated as a scalar field and it is assumed that the source is not spatially coherent, so light vibrations from its different regions are statistically independent.

The measurements are confined to a specific plane that radio astronomers call the *u-v plane*. The variables u and v are two coordinates of the projected aperture separation. If the unit vector $\hat{\mathbf{k}}$ points from the interferometer to the source, and the unit vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ lie in a plane tangential to the celestial sphere at the source, then we can write

$$u = \frac{\hat{\mathbf{i}} \cdot (\mathbf{p}_2 - \mathbf{p}_1)}{\lambda}, \quad \text{and} \quad v = \frac{\hat{\mathbf{j}} \cdot (\mathbf{p}_2 - \mathbf{p}_1)}{\lambda},$$

with

$$\tau = \frac{\hat{\mathbf{k}} \cdot (\mathbf{p}_2 - \mathbf{p}_1)}{c}.$$

Here the propagation time τ replaces the variable w that is normally used in radio astronomy to describe the third coordinate of the projected aperture separation. The *u-v plane* is located where $\tau = 0$.

Coordinates within the source may be described by the direction cosines l and m , as measured with respect to the vector $\hat{\mathbf{k}}$ from the interferometer. The brightness distribution of the source is then described by the function $I(l, m)$. The theorem can

be written as follows:

$$\gamma_{12}(0) = \frac{e^{-i2\pi w} \int_{\kappa} \int I(l, m) e^{-i2\pi[ul+vm]} dl dm}{\int_{\kappa} \int I(l, m) dl dm}, \quad (1.1)$$

where $\gamma_{12}(0)$ is the complex degree of coherence of the source describing the fringe visibility and phase. This is in the form of a normalised two-dimensional Fourier transform. The brightness distribution $I(l, m)$ is related to measurements of the fringes at different values of u and v determined by the aperture separations, or baselines. Therefore if the fringes are observed at different locations in the u - v plane it is possible to reconstruct the source through an inverse transform.

1.1.2 Path compensation and fringe tracking

In general τ would be non-zero, because the points \mathbf{p}_1 and \mathbf{p}_2 are fixed to the ground and rotate with the earth. However, the pathlengths can be compensated by adjusting s_1 and s_2 within the interferometer. We can redefine τ so that

$$\tau = \frac{\hat{\mathbf{k}} \cdot (\mathbf{p}_2 - \mathbf{p}_1)}{c} + \frac{(s_2 - s_1)}{c}, \quad (1.2)$$

and choose s_1 and s_2 so that $\tau = 0$. The path-difference $s_2 - s_1$ must then be constantly changing to track the vector $\hat{\mathbf{k}}$ that follows the source across the sky. The standard approach in optical interferometry is to track the delay using a moving optical element, and to monitor its motions with laser metrology.

However, pathlength compensation is further complicated by the effects of the turbulent atmosphere. The wavefronts of starlight are distorted by changes in the index of refraction of dry air, and path fluctuations are induced that can be many wavelengths long. Moreover, the changes evolve on timescales of milliseconds or tens of milliseconds. Thompson et al. (1986) point out that in optical interferometry it is therefore more difficult to obtain an accurate phase calibration, and in many cases only the visibility amplitude is measured. The severity of the wavefront distortions is often quoted in terms of Fried's coherence length, r_0 , which is the separation of two points for which the rms phase difference is 2.6 radians. At radio wavelengths r_0 is many kilometers wide, but in the optical it can be smaller than 10 cm.

With large optical telescopes it is common for the wave distortions to be smaller than the aperture size. When the apertures are greater than r_0 there are numerous wavefront tilts in the aperture, and these can each give rise to separate images in the focal plane. An image of a star is not diffraction limited, but is distorted and broken into many fragments. These appear as speckles, similar to the speckles observed in diffuse laser light (Labeyrie, 1970; Dainty, 1984), and as the phase structure evolves

with time the speckles appear to boil and change shape. Examples of this are shown in the review by Labeyrie (1978). If the light from pairs of large apertures is coherently combined then it would be difficult to calibrate for the effects of atmospheric phase.

It is for this reason that modern Michelson stellar interferometers use small r_0 sized apertures and active control of wavefront tilt. The small apertures ensure that the wavefronts passing into the instrument are essentially flat, even though they may be tilted with respect to each other. The light through each aperture forms a diffraction limited image of the star, and the tilts can be stabilised by monitoring the image motion. This information is used to control piezo mirrors which are servoed so that the beams are combined with zero tilt and shear. If the combined light was then focused to an image plane then the two stellar images would appear to be stationary and perfectly superimposed. However, the differential path fluctuations would still exist.

Residual path-differences can seriously affect the measured visibility. The bandwidth that is used for an observation determines the largest path-difference over which fringes will be visible, also termed the coherence length. The visibility of the fringes drops at increasing values of τ , and falls to zero when the path difference is equal to the coherence length. The atmosphere is capable of inducing rapid pathlength changes, and can therefore cause the fringe visibility to change over short timescales.

Fringe tracking and path compensation are important aspects of stellar interferometry. If the path errors cannot be corrected then the only way to reduce the visibility losses is to narrow the bandwidth and thereby increase the coherence length. However, large bandwidths are clearly more desirable; they permit more light to be used in the measurement and thus improve the sensitivity to faint objects.

1.2 The Sydney University Stellar Interferometer

SUSI is a modern Michelson stellar interferometer, located at the site of the Australia Telescope near Narrabri, New South Wales. It is designed to measure the angular diameters of stars to an accuracy of $\pm 2\%$ with a limiting magnitude of $+7.5$ and a resolution of $7.5 \cdot 10^{-5}$ arc seconds (Davis, 1985b; Davis et al. 1992).

The layout of the interferometer is shown in Fig. 1.1. It has 12 siderostats on a north-south baseline with possible aperture separations in a geometrical progression from 5 to 640 m. Small apertures and active wavefront tilt correction are used to minimise the coherence loss caused by atmospheric turbulence. Light is steered into the main building via relay mirrors which guide it through an evacuated pipe. On entering the building the light leaves the pipe and emerges into a temperature controlled enclosure. There it passes through a beam reducing telescope, which compresses the beam diameter by a factor of three, and then along an optical pathlength compensator

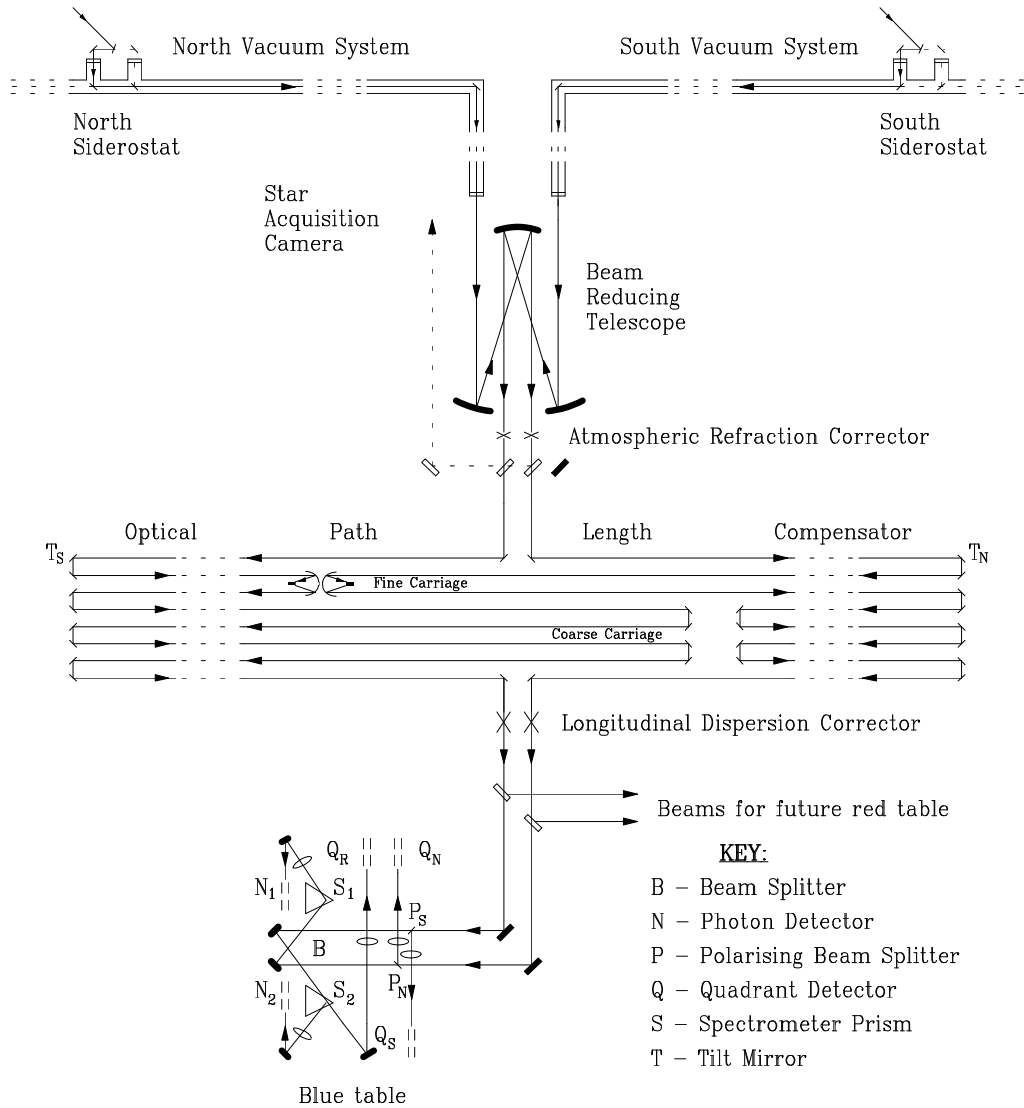


Figure 1.1: The layout of the Sydney University Stellar Interferometer. Light arrives at the north and south siderostats and is redirected into the main building. It passes through the Beam Reducing Telescope, compressing the beam diameter to about 4 cm, and then is sent back and forth along the Optical Path Length Compensator before arriving at the beam-combining table. During this process the effects of longitudinal dispersion, atmospheric refraction, and atmospheric tilt, are compensated. (Drawing courtesy of Theo ten Brummelaar.)

which moves so that the coherence envelope of the starlight remains stationary, despite the changing altitude and azimuth of the star. The light is brought to the main optical table, combined, and then dispersed by a prism spectrometer. Photomultiplier tubes are used to count photons within a bandwidth of ~ 1 nm, and the fringe visibility is measured using the counts in each of the two combined beams, as described by Tango and Twiss (1980).

For this to be successful, all components of the interferometer must be free of vibration. This is true not only for the siderostats, which have foundations that extend several meters into bedrock, but also for the path compensating carriage, which must move during the course of an observation.

1.3 The Optical Pathlength Compensator

The optical pathlength compensator has been outlined previously by Booth et al. (1992), but will be reviewed here briefly. It has been designed for two carriages mounted on rails on a 70 m long reinforced concrete pier: a coarse carriage that remains stationary during an observation, and a fine carriage that is microstepped under servo control. The path compensation is done in air in a temperature controlled environment with provision for the correction of longitudinal dispersion.

There are two identical laser metrology systems which independently monitor the position of each side of the fine compensation carriage, one system at the northern and southern end of the path compensator. HeNe lasers are used with an operating wavelength of $1.15 \mu\text{m}$, and a collimated beam diameter of 30 mm. Although the beams from the metrology lasers travel the same pathlengths as the starlight, they are separated from it both spatially and in wavelength. The metrology system is able to track movement errors to an accuracy of 18 nm in optical path difference.

Large scale path compensation is done differentially: the fine carriage contains two catseye reflectors which are pointed in opposite directions along the rails. When the fine carriage is displaced by a distance x_f it will remove a pathlength $2x_f$ from one arm whilst adding it to the other, producing a total pathlength change of $4x_f$. The total range of adjustment for the fine carriage is therefore 280 m (4×70 m), or ± 140 m. The smallest pathlength adjustments are done with piezo electric mirrors mounted on the carriage itself. At the focus of each catseye there is a piezo controlled mirror, one with a throw of $30 \mu\text{m}$ and the other with a throw of $15 \mu\text{m}$. Each piezo acts in only one arm of the interferometer and produces a path adjustment of only twice its displacement, $2x_{p1}$ and $2x_{p2}$. The expression for the total path adjustment is therefore

$$4(x_f - x_0) - 2(x_{p1} - x_{p2}), \quad (1.3)$$

where x_0 is the white-light fringe position of the carriage for autocollimation from the

siderostats. This expression indicates that a positive motion of a piezo shortens the pathlength in that arm and that if they both move by equal amounts the net effect is zero pathlength change.

1.4 Motivation for this Work

In its design SUSI was intended to operate using fringe tracking so that larger bandwidths could be used. The 11.4 m prototype interferometer, upon which the instrument was based, had not used fringe tracking, but instead relied on measurements of the coherence envelope of the fringes (Davis & Tango, 1986). As will be discussed in Chapt. 2, this method is time consuming and inefficient, but nevertheless useful for observations of bright stars at short baselines. For observations of faint objects at long baselines, fringe tracking was deemed to be essential, and was therefore included in the design.

Phase tracking, however, was not considered to be important. The proposed scientific program was primarily concerned with the measurement of the angular diameters of stars and the separation of binary stars, which are all objects that possess a simple symmetry and can be described by visibility amplitudes alone—without the knowledge of visibility phase. Phase tracking was therefore not required.

Group Delay Tracking

Group delay tracking was suggested as a suitable alternative. The theoretical and experimental work in this subject form the bulk of this thesis. By this technique the starlight is brought to the beam combiner, dispersed through a spectrometer, and imaged onto a photon counting array detector. Whenever the path-difference between the two beams is greater than zero there will be fringes visible in the spectrum, with the number of fringes directly proportional to the path difference. Thus by determining a fringe frequency it is possible to measure the path-difference in real-time and correct for errors induced by atmospheric turbulence. Tango and Twiss (1980) make reference to this, calling it a form of AFT (automatic fringe tracking) and attributing the idea to W.H. Steel.

Photon counting array detectors

At SUSI it had also been envisaged that photon counting array detectors would eventually replace the photomultipliers that are presently used. Measurements of visibility could then be done at many different wavelengths simultaneously. The wavelengths of observation would be determined by the location of the pixels in the array, and the

bandwidths would be set by the pixel size or the width of pixel groupings.

Furthermore, the same information could be used for group delay tracking. The development of an array detector was therefore considered an important aspect of the design of the fringe tracking system, and arrangements were made with Harvard University for the purchase of a PAPA camera. Although unforeseen at the time, the detector proved to have defects which made it unsuited for observational work. The redesign and construction of a new PAPA camera became a major component of this research.

1.5 Outline of the Thesis

The thesis has been divided into four parts: an introduction and review; the design of the PAPA camera; the theory of group delay tracking; and the presentation and discussion of experimental work.

Part I includes this chapter, and will be completed with a review of fringe tracking methods, which follows in Chapt. 2. Only methods which use wavefront-tilt compensation are considered in the review. Previous investigations of group delay tracking are discussed, and the different approaches to the data analysis are presented.

Part II is devoted to the description of the PAPA camera which was built as part of this work. Both practical and theoretical aspects of the design of PAPA cameras are presented in Chapt. 3. Included there is a review of previous designs, a discussion of optical tolerances, and a description of image artifacts. The rebuilt detector is described in Chapt. 4, where the design choices are discussed and the camera's performance is presented.

Part III describes the theory and limitations of using the Fast Fourier Transform for group delay tracking. The discussion is introduced in Chapt. 5, where the approach to the subject is outlined. The influence of the detector on the measurement, including the resolution and the response to large delays, is shown in Chapt. 6. Signal-to-noise issues and the probability of tracking error are presented in Chapt. 7 where the low light level limitations are discussed. The effects of atmospheric turbulence are then reviewed in Chapt. 8 and its influence on the tracking is predicted.

Part IV discusses the experiments which were performed to test group delay tracking. Chapter 9 describes the hardware that was built to calculate a power spectrum from the photon data, and also details the signal processing software that is used. The laboratory experiments and the stellar observations are then presented in Chaps. 10 and 11.

Chapter 2

Fringe Detection in Stellar Interferometry

2.1 Fringe Tracking with Wavefront Tilt Compensation

Methods of fringe detection and tracking will now be reviewed, with the discussion limited to those methods that use small apertures and wavefront tilt compensation. It is perhaps surprising that even without tilt correction it is still possible to record fringes. However, a two-dimensional detector must be used in either the image plane or the pupil plane of the combined beams. The interpretation of the fringes is then more difficult, but is facilitated if the light is dispersed, because the slope of the fringes as a function of wavelength yields information about the optical path difference. Examples of this are discussed by Vakili et al. (1988), Rabbia (1988), and Schumacher et al. (1992).

When apertures smaller than r_0 are used the interfering wavefronts are essentially flat, and with tilt correction they can be made parallel; the optical path differences then depend on only one parameter—the difference in arrival times of the two plane waves. Interferometers which have used tilt correction include the prototype at Monteporzio (Tango, 1979), the 11.4 m Sydney University Prototype (Davis & Tango, 1985), the Mark III interferometer (Clark et al., 1986), the Cambridge Optical Aperture Synthesis Telescope (Cox, 1992), and the Sydney University Stellar Interferometer (ten Brummelaar, 1992). Tilt correction is considered an essential part of most modern stellar interferometers.

After a brief outline of some background theory, the coherence envelope is discussed followed by a description of delay curve measurements. Methods of active fringe tracking are described, and group delay tracking is introduced. The results of previous work in this field are then summarised.

2.2 The Interference of Quasi-Monochromatic Light

The expression for the interference of quasi-monochromatic light has been derived by many authors, see for instance Hecht (1987). The intensity of the combined light can be expressed as

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} |\gamma_{12}| \cos(\alpha_{12} - \varphi),$$

where

$$\varphi = \frac{2\pi}{\lambda}(s_2 - s_1).$$

λ is the wavelength of the light, I_1 and I_2 are the intensities of the light in each arm of the interferometer, and γ_{12} is the complex degree of coherence with modulus $|\gamma_{12}|$ and argument $\alpha_{12} - \varphi$, where φ arises from the path difference, and α_{12} contains information about the source. The pathlengths s_1 and s_2 are the same as described previously in §1.1. The contrast, or visibility, of the fringes is the ratio of the fringe amplitude to the total background illumination,

$$\mathcal{V} = \frac{2\sqrt{I_1 I_2} |\gamma_{12}|}{I_1 + I_2}.$$

If we make a change of variables introducing the spectroscopic wavenumber $\kappa = 1/\lambda$, letting

$$I_s = 2\sqrt{I_1 I_2}, \quad I_b = I_1 + I_2 - I_s,$$

and

$$x = (s_2 - s_1),$$

then we have

$$I(\kappa, x) = I_s \left[1 + |\gamma_{12}| \cos(2\pi\kappa x - \alpha_{12}) \right] + I_b. \quad (2.1)$$

The substitution of $x = s_2 - s_1$ indicates that tilt correction is being used in the combined beams. Under conditions where $I_1 = I_2$ then the visibility of the fringes is the modulus of the complex degree of coherence,

$$\mathcal{V} = |\gamma_{12}|.$$

A source that is *not* quasi-monochromatic may still be treated as such if it is observed with an instrumental bandwidth that is sufficiently small. Bright fringes will occur wherever the path difference x is an integer multiple of 2π at most wavelengths. This will be so at *all* wavelengths only when x is zero and when the dispersion is the same in each arm of the interferometer. The reduction in the fringe visibility at increasing values of x is described by the coherence envelope.

2.3 The Coherence Envelope

All interferometers observe fringes over a finite bandwidth, and the recorded intensity is the integral of $I(\kappa, x)$, over wavenumber, weighted by a filter function $W(\kappa)$.

$$I(\bar{\kappa}, x) = \int_{-\infty}^{\infty} W(\kappa - \bar{\kappa}) I(\kappa, x) d\kappa, \quad (2.2)$$

where $\bar{\kappa}$ is the center of the passband.

The filter function includes both the shape of the bandpass and the frequency response of the detector; it has values that are large within the bandwidth, and near zero outside. The result of this averaging is to reduce the sensitivity of the interferometer to fringes of large delay: when the bandwidth $\Delta\kappa$ partially spans a fringe (in the wavenumber domain κ) then the visibility appears to be reduced. This is simple to illustrate.

Let us introduce a change of variables, such that $\kappa' = \kappa - \bar{\kappa}$, and perform the integration in Eq. 2.2 with respect to κ' . If we insert Eq. 2.1 into Eq. 2.2 and rearrange the terms we have

$$\begin{aligned} I(\bar{\kappa}, x) = I_s [& 1 + |\gamma| \cos(2\pi\bar{\kappa}x - \alpha) \int_{-\infty}^{\infty} W(\kappa') \cos(2\pi\kappa'x) d\kappa' \\ & - |\gamma| \sin(2\pi\bar{\kappa}x - \alpha) \int_{-\infty}^{\infty} W(\kappa') \sin(2\pi\kappa'x) d\kappa'] \end{aligned} \quad (2.3)$$

where the subscripts have been dropped from γ_{12} and α_{12} . Now if we define $\Omega(x)$ as the Fourier transform of $W(\kappa)$, then we have:

$$\Omega(x) = |\Omega(x)| e^{j\phi_\Omega} = \int_{-\infty}^{\infty} W(\kappa) e^{j2\pi\kappa x} d\kappa,$$

and therefore

$$\begin{aligned} |\Omega(x)| \cos \phi_\Omega &= \int_{-\infty}^{\infty} W(\kappa) \cos(2\pi\kappa x) d\kappa, \\ |\Omega(x)| \sin \phi_\Omega &= \int_{-\infty}^{\infty} W(\kappa) \sin(2\pi\kappa x) d\kappa. \end{aligned}$$

Equation 2.2 may therefore be written in the form

$$I(\bar{\kappa}, x) = I_s \left[1 + |\gamma_x| \cos(2\pi\bar{\kappa}x - \alpha + \phi_\Omega) \right] \quad (2.4)$$

where the apparent visibility $|\gamma_x|$ is the product of the true visibility and the modulus of the Fourier transform of the filter function, evaluated at the current delay:

$$|\gamma_x| = |\gamma| |\Omega(x)|.$$

The transfer function $\Omega(x)$ describes the *coherence envelope*. If $W(\kappa)$ is symmetric then $\Omega(x)$ is real valued, $\phi_\Omega = 0$, and only at zero delay, where the envelope is at its peak, is the true visibility observed.

The above calculation is more difficult if there is unequal dispersion in the two arms of the interferometer, or if the spectrum of the source has structure within the passband. The central fringe will then be displaced and the overall visibility of the fringes will be reduced, because there will no longer be a single delay for which constructive interference occurs at all wavelengths (Tango, 1990).

A Rectangular Bandpass

If a detector has a rectangular bandpass then its coherence envelope would resemble a sinc function:

$$W(\kappa) = \begin{cases} 0, & |\kappa| > \Delta\kappa/2 \\ 1, & |\kappa| < \Delta\kappa/2 \end{cases} \quad \text{and} \quad \Omega(x) = |\Delta\kappa| \frac{\sin \pi x \Delta\kappa}{\pi x \Delta\kappa}.$$

If a bandwidth of $\Delta\lambda$ is used at a wavelength λ , then the same interval expressed in wavenumber is as follows:

$$\Delta\kappa = \frac{1}{(\lambda - \Delta\lambda/2)} - \frac{1}{(\lambda + \Delta\lambda/2)}, \quad \text{therefore} \quad \Delta\kappa = \frac{\Delta\lambda}{\lambda^2 - (\Delta\lambda/2)^2}.$$

If we assume that the fractional bandwidth is very small then we can ignore the second term in the denominator.

$$\Delta\kappa = \frac{\Delta\lambda}{\lambda^2}.$$

The sinc function is characterised by the location of its first zero crossing, where $x = 1/\Delta\kappa$. This distance can be thought of as the coherence length of the starlight under observation. If, for instance, a bandwidth is 0.4 nm is used at a wavelength of 450nm, then the coherence effects will be seen when the delay offset is less than $\sim 500\mu\text{m}$. The observed visibility will increase as the delay approaches zero, decrease as zero delay is passed, and eventually approach the level of the noise when the delay exceeds the coherence length.

2.4 Measurements of Delay Curves

Tango & Twiss (1980) describe the estimation of visibility based on measurements of the coherence envelope. This technique has been used at the Monteporzio interferometer (Tango, 1979) and also at the Sydney University 11.4 m Prototype (Davis & Tango, 1985). A photomultiplier is used to count the number of photons received in each of the combined beams, and observations are gated to timescales where the atmosphere is stable. The recorded photon counts in each channel, A and B , are proportional to the intensity of the light received there.

$$\begin{aligned} A &\propto I_s[1 + |\gamma_x| \cos(2\pi\kappa x - \alpha)], \\ B &\propto I_s[1 - |\gamma_x| \cos(2\pi\kappa x - \alpha)], \end{aligned}$$

where $|\gamma_x|$ is the apparent visibility, and the difference in sign is brought about by the optical beam combiner. The total light in the system is always $2I_s$: if fringes are present then one beam becomes brighter as the other becomes dimmer.

A visibility estimate may be derived by forming the ratio of the squares of the sum and difference terms of A and B .

$$\left[|\gamma_x| \cos(2\pi\kappa x - \alpha) \right]^2 = \frac{(A - B)^2}{(A + B)^2}.$$

The phase of the fringes is embedded in the measurement and must be removed if the magnitude of the visibility is to be found.

In the absence of fringe tracking the pathlengths in each arm will vary according to the atmospheric turbulence. Consequently, the phase-difference will wander about the mean open-loop tracking position, and will change from one sample time to another. However, when narrow bandwidths are used these phase variations, of 10 wavelengths or more, represent only a small fraction of the coherence length, and so despite the fringe wander the location on the envelope will remain roughly constant. The apparent visibility is then measured by averaging a large number of samples to eliminate the phase term. The delay-line can then be stepped to move through the coherence envelope, with measurements of the visibility $|\gamma_x|$ made at each location. This produces what Davis & Tango (1986) have called a *delay curve*. If the shape of the bandpass is known then a model of the coherence envelope can be fitted to the data allowing the peak value, $|\gamma|$, to be interpolated.

At zero delay the observed visibility of an unresolved star ought to be unity, but in practice that is never the case, because phase aberrations are introduced from several sources. The optical surfaces within the interferometer cannot be perfectly flat, the incoming light must pass through a series of diffracting apertures, the vacuum windows and beam-splitters may be non-homogeneous, and there will always be small atmospheric phase errors, even after tilt-correction. The visibility measurements must

therefore be calibrated using observations of unresolved stars. Davis & Tango (1986) used this approach to measure the angular diameter of α CMa with β CMa as a calibrator.

Limitations of Envelope Tracking

The major drawback of envelope observations, or *envelope tracking* as it is termed, is the low signal-to-noise ratio. Narrow bandwidths, typically 0.5–1.5 nm, are necessary for this method to work, and consequently few photons arrive in a sampling interval. A large number of samples need to be integrated to reduce the noise, and for even the brightest stars an observation time greater than 15 minutes is required to record the delay curve. The expected signal-to-noise ratio for this mode of operation is discussed by Tango & Twiss (1980).

If the baseline of the interferometer is not precisely known then there will be a systematic error in the rate of delay tracking, causing the coherence envelope to move. As a result the measured envelope will appear expanded or contracted, dependent upon the direction of the tracking error. If the baseline estimate is poor then the fringe envelope may drift so quickly that fringes cannot be detected. However, once fringes are observed, the tracking can be iteratively improved by monitoring the motions of the envelope. Adjustment are made to the baseline estimate (and therefore the tracking velocity) until the envelope remains stationary. All changes must be performed through human intervention, and this method is therefore time consuming and inefficient.

2.5 Pathlength Modulation

There are many ways to determine the phase of a wavefront, but all require a time-dependent phase modulation (Creath, 1988). One technique, described by Wyant (1975) and related to methods in Fourier spectroscopy (Connes, 1970), has been successfully used in the Mark III interferometer (Shao & Staelin, 1977; Shao & Staelin, 1980; and Shao et al., 1988). In that instrument the red portion of the spectrum, ~ 650 – 900 nm, is used for white-light fringe tracking, and three selected spectral-bands, each 20 nm wide, are used for data collection; these are set by interference filters and are typically located at 800, 550, and 450 nm (Mozurkewich et al., 1991). Although the bandwidth of the fringe tracker has a coherence length of only $1.8 \mu\text{m}$, the coherence length for the data collection is between 8 and $25 \mu\text{m}$. Data is only collected when the fringe tracker is locked onto the fringe.

In this technique the pathlength in one arm of the interferometer is modulated

so that $x = \bar{x} + x_n$, and the fringe then takes on the form

$$I(\kappa, x) = I_s \left[1 + |\gamma_x| \cos \left(2\pi\kappa(\bar{x} + x_n) - \alpha \right) \right],$$

where

$$x_n = \bar{\lambda} \frac{n}{4}, \quad n = 0, 1, 2, 3.$$

In practice the pathlength is not stepped, but is varied as a triangle wave with a peak-to-peak amplitude of one wavelength, $\bar{\lambda} = 800$ nm. An example of integrating across a waveform is given by Wyant (1975) and the principle is discussed by Creath (1988) who calls this the ‘four bucket’ technique. During the course of one cycle, the photon counts A , B , C , and D are recorded,

$$\begin{aligned} A &\propto I_s \left[1 + |\gamma_x| \cos(2\pi\kappa\bar{x} - \alpha) \right], \\ B &\propto I_s \left[1 - |\gamma_x| \sin(2\pi\kappa\bar{x} - \alpha) \right], \\ C &\propto I_s \left[1 - |\gamma_x| \cos(2\pi\kappa\bar{x} - \alpha) \right], \\ D &\propto I_s \left[1 + |\gamma_x| \sin(2\pi\kappa\bar{x} - \alpha) \right], \end{aligned}$$

so that

$$2\pi\kappa\bar{x} - \alpha = \tan^{-1} \left(\frac{A - C}{D - B} \right),$$

and

$$|\gamma_x|^2 \propto \frac{(A - C)^2 + (D - B)^2}{A + B + C + D}.$$

The pathlength \bar{x} can therefore be determined and fringe phase can be tracked (Shao & Staelin, 1980). Although a 2π ambiguity exists in the measurement, there are phase ‘unwrapping’ procedures which can be used to track large path changes (Colavita et al., 1987). The visibility measurements at each wavelength are calculated by a different equation which subtracts the dark counts and removes the photon-noise bias (Shao et al., 1988).

There are obvious advantages to this approach as compared with envelope tracking. The most important is that wide bandwidths are used for fringe location. The Mark III uses a photon-counting avalanche photodiode, a bandwidth of 650–900 nm, and apertures 7.5 cm in diameter. This yields a magnitude limit of ~ 4.5 for its fringe detection (Mozurkewich et al., 1991; Armstrong et al., 1992). Envelope measurement, however, are limited to bright stars at short baselines.

The paths are modulated for a wavelength of 800 nm, and for observations at the shorter wavelengths, at 550 and 450 nm, it is necessary to rebin the data. It must appear to the data processor that the modulation is one wavelength peak-to-peak, simultaneously at each of the three different wavelengths. A dead-time is therefore inserted in the data processing around the peaks of the triangle wave (Mozurkewich et

al., 1991). Data points are ignored when path excursions are greater than the desired wavelength, and boundaries of the *A*, *B*, *C*, and *D* bins are re-defined. Data is thrown away and, of course, this reduces the sensitivity.

The new optical table of the IT2 interferometer (B. Sorrente et al., 1992) also uses path modulation, but within a sub-interferometer that is separate from the main instrument. Measurements of pathlength therefore do not interfere with visibility measurements. This is an interesting approach, since it avoids the problem discussed in the previous paragraph. However, path equality in the sub-interferometer would not imply the same in the main instrument, since there would be added pathlengths involved.

2.6 Channeled Spectra and Group Delay Tracking

There is only one other method that has been discussed in the literature for delay tracking using small apertures with wavefront-tilt correction. This has become known as Group Delay Tracking, and relies on the observation of fringes in dispersed stellar spectra. The spectrum of the combined starlight will be modulated by fringes whenever the pathlength difference between the two arms of the interferometer is non-zero. These have been termed Edser–Butler fringes, or fringes of equal chromatic order, and produce a *channeled* spectrum (Steele, 1987). If the aperture pupils were superimposed and the wavefronts were parallel when combined, then the bright bands of the fringes will be perpendicular to the direction of dispersion. This is expressed by the wavelength dependence contained in Eq. 2.1. Dispersing the light allows the fringes to be seen even when the path difference is several hundred wavelengths (Michelson & Pease, 1921). Steel (1987) states that

Fringes of equal chromatic order provide the most sensitive method of adjusting the interferometer delay to obtain maximum visibility with a source of finite spectral bandwidth. It is an old technique... used by Michelson & Pease (1921), and frequently rediscovered since.

The re-discoveries include work published in the last five years (Smith & Dobson, 1989). The same principles have been used previously to study the thickness of thin films, as discussed by Born & Wolf (1986). The number of fringes in the spectrum, n , is directly proportional to the pathlength error: reducing the fringe frequency brings the delay closer to zero. It can be seen from Eq. 2.1 that

$$n = \frac{x}{\lambda_{min}} - \frac{x}{\lambda_{max}}.$$

All that is needed to track the delay, x , is to determine the number of fringes contained in the spectrum, or the frequency of the fringes.

The idea of applying this technique to modern stellar interferometry has been attributed to Steel by Tango & Twiss (1980), and also to Labeyrie by Koechlin (1985). There have been several papers in conference proceedings which have discussed the probable limitations of the method. These include simulations for the IOTA project performed by Nisenson & Traub (1987) and Traub (1990a); simulations for the COAST interferometer performed by Buscher (1989); and signal-to-noise predictions for ‘photon-starved’ operation with the Mark III interferometer (Shao et al., 1988). It has also been reviewed along with other methods applicable to space-based optical interferometry (Shao & Colavita, 1992). But to the author’s knowledge the only experiments using starlight were performed by Kim (1989) as part of his Masters thesis at the Massachusetts Institute of Technology, in conjunction with the Mark III group. The different approaches to the data processing will now be reviewed.

2.6.1 Model fitting by least-squares

Traub (1990a) describes a method which uses a cross-correlation of the data with model functions. It is assumed that a family of functions exist which will closely fit the data providing certain parameters are adjusted. For example a model may be written

$$f_k(x) = I_k [1 + |\gamma_k| \cos(2\pi\kappa_k x + \phi_k)],$$

which accounts for the shape of the stellar spectrum I_k , a varying visibility of the fringes $|\gamma_k|$, non-linearities in the detector κ_k , and fringe spacings for arbitrary delays x . It follows that if all these parameters are chosen correctly then it will minimise the least-squared difference between the data and the model. If we were to consider the delay by itself then we would perform the minimisation by taking the partial derivative with respect to x of the mean squared difference, equating it to zero, and solving for the delay. If the data are represented by the set g_k then we have

$$\frac{\partial}{\partial x} \left[\sum_{k=1}^K [g_k - f_k(x)]^2 \right] = 0,$$

which can be written in full as

$$\frac{\partial}{\partial x} \left[\sum_{k=1}^K [g_k^2 - 2g_k f_k(x) + f_k^2(x)] \right] = 0.$$

In this equation only the cross term is of interest. The sum of the g_k^2 terms is a constant and contributes nothing to the minimization. Furthermore, if the model $f_k(x)$ was normalised correctly then the sum of the $f_k^2(x)$ terms would be independent of x , and therefore would also be a constant. We can now express the minimization of the mean square difference as

$$\frac{\partial}{\partial x} \left[\sum_{k=1}^K g_k f_k(x) \right] = 0,$$

where the function $f_k(x)$ maximizes the sum of the cross terms. The quantity in brackets is simply the cross-correlation between the model and the data, calculated at zero lag. Traub (1990a) presents simulations of pathlength motions with peak-to-valley excursions of $1.3 \mu\text{m}$ over 1 second with $|\gamma| = 1.0$. He concludes that delay tracking should be possible at count rates as low as 10 photons per coherence time, with a position uncertainty of $\sim 0.2\lambda$.

2.6.2 Optimal and Kalman filtering

It should be noted that the statistics of the noise are not accounted for using a cross-correlation, and it is therefore not an optimal filter. The concept of optimal, or Wiener filtering, requires knowledge of the noise process: the power spectra of the noise and signal are used to determine the shape of the filter. If the noise is absent then no filtering occurs, but if the noise dominates then most of the detected power is rejected by the filter (Press et al., 1992, §13.3).

The Kalman filter is a *recursive* linear mean-squared estimator (Kalman & Bucy, 1961) and has been described in several texts, including Bryson & Ho (1969) and Boziz (1979). A volume of selected reprints describing Kalman filters has been published by IEEE Press (Sorenson, 1985). It is historically important because the recursion permitted a tremendous reduction in computational requirements, and it was expressed in a framework that had a unifying influence on previous research (Sorenson, 1970). Although the filters are designed to estimate parameters that have a linear relationship with the measured data, they have seen their greatest application in non-linear systems. This has included the determination of aircraft trajectories, spacecraft orbits, power station control, and the demographics of cattle production (Sorenson, 1983). Non-linear processes must be ‘linearized’ by approximating their behavior with linear equations, at least in the neighborhood of the current estimate, and this may include the derivatives of parameters as well. If the approximation errors become large then the solution will diverge.

Kalman filters were mentioned by Nisenson & Traub (1987) in connection with group delay tracking, as a possible means of improving the performance of power spectrum analysis. Reasenberg (1990) used a Kalman filter in simulations to evaluate their use with the IOTA interferometer. However, the performance of his filter is difficult to judge based on his presentation, which is restricted to a discussion of the useful model parameters. He does not discuss the low light level limitations, but observes that filters which include estimates of the rate-of-change of pathlength would be more accurate than ones which estimate pathlength difference alone.

2.6.3 Spectrum analysis and the Fast Fourier Transform

The frequency of the fringes in the channeled spectrum can be determined using methods of power spectrum analysis. One of the simplest and quickest is the Fast Fourier Transform (FFT). It will be compared to other methods.

Modern Spectrum Analysis

Approaches to spectrum analysis are frequently based on the use of the FFT. Although it is computationally efficient, it suffers from several drawbacks—notably the poor resolution, aliasing, and spectral leakage (discussed in Appendix F). The review of modern spectrum analysis by Kay & Marple (1981) therefore begins with a discussion of the FFT, and proceeds to use it as a touchstone for comparing other methods. Each new technique has developed based on a different model for the signal and uses different parameters to characterize the underlying process.

The utility of a particular technique depends on how well its assumptions model the data. For instance, the FFT assumes that the data represent a series of harmonically related sinusoids. If there exists one sinusoid that does not coincide with any of the harmonics, then its corresponding power spectrum will be poorly reconstructed. Likewise, if the data were sampled at irregular intervals then the spectrum will be distorted. Advances in spectrum analysis have come from deriving power spectra from more accurate assumptions. A selection of the different models is given in Table III of the review by Kay & Marple (1981). These include models which contain a sum of non-harmonically related sinusoids (Prony spectral line decomposition, and Pisarenko harmonic decomposition), non-harmonically related damped exponentials (Extended Prony Method), autoregressive or all-pole processes (AR and Maximum Entropy Method), moving average or all-zero processes (MA), and autoregressive moving average or pole-zero processes (ARMA). Techniques are also available that will process data that were sampled on an irregular grid, or that have samples missing (Press et al., 1992 §13.8). This would be common in records of channeled spectra when there are fewer photons per frame than detector pixels. In the Lomb–Scargle method the missing samples do not contribute to the calculation of the power spectrum; by contrast the FFT treats missing samples as zero-valued and gives them equal weight in the calculation (Press & Rybicki, 1989). Spectacular improvements are possible if the proper model is chosen and a strong signal is present.

Unfortunately, these methods will fail when they are asked to derive spectra from processes that deviate from their model. This can occur for AR methods simply by adding observation noise to the data: at low signal-to-noise levels the resolution is no better than an FFT approach (Kay & Marple, 1981). Moreover, most of the methods are ill adapted for real-time processing. While it may be possible to determine

the parameters that describe the spectrum, one must then recalculate the spectrum numerous times to find the peak, performing lengthy summations. Even the ‘fast’ version of the Lomb–Scargle method, although it uses an FFT, is not a real-time method.

The Fast Fourier Transform

The characteristics of the FFT and its relationship to the integral Fourier transform, are described in Appendix F. It is identical to the Discrete Fourier Transform (DFT), but bears a different name because of the algorithm used to implement it. The FFT is an obvious candidate for real-time applications, as there exist integrated circuits which will calculate long FFTs in less than a millisecond, fast enough to track phase-difference changes in atmospheric turbulence.

There are two basic ways of using an FFT to compute power spectra: the first is the Blackman–Tukey algorithm, and the second is the periodogram (Kay & Marple, 1981). The Blackman–Tukey method is a discrete time implementation of the Wiener–Khinchin theorem; the autocorrelation of the data is formed and then Fourier transformed to arrive at the power spectrum. It was the most popular form of power spectrum estimation prior to the introduction of the FFT. The periodogram estimate is a more direct approach of applying a discrete Fourier transform to the data and calculating the mean squared value of the magnitude of each frequency component. The two methods are not equivalent, however, under most circumstances they will yield identical results and have the same properties.

Nisenson & Traub (1987) ran simulations using the periodogram and concluded that tracking would be possible with as few as 3 photons per frame. The simulations included path-difference changes in the form of a random-walk, fringe visibility of $|\gamma| = 1.0$, and an integration of 100 frames, each 10 ms long, to represent one second’s worth of data. Buscher (1988; 1989) performed similar work, this time using path motions which were derived by the Kolmogorov–Taylor approximation (see Chapt. 8). Much larger pathlength changes were modelled and a low-pass filter was used with a time constant of 75 frames. The fringe visibility was chosen to be $|\gamma| = 0.25$ and simulations were run for 4, 8, 16, and 32 photons-per-frame. The results indicated that tracking would be possible for between 8 and 16 photons-per-frame.

Colavita and Shao (1988) described the signal-to-noise characteristics of group delay tracking for ‘photon-starved’ conditions, where there would be too little light for active fringe tracking. A similar discussion appeared in their paper describing the Mark III interferometer (Shao et al., 1988). They suggested that the limiting magnitude would be set by an observation time of 4000 seconds, just over an hour, and presented a simulation with 2.5 photons per frame. Furthermore, they announced their intention to test group delay tracking at the Mark III. This was to be using the PAPA

camera they have for tilt-correction (Clark, 1986) also described by Gonsiorowski (1986), and processing the data with the Blackman–Tukey algorithm.

Those tests were the basis for the Masters thesis by Kim (1989). He observed the stars α Lyr, η Cyg, δ And, and α And, on the nights of 22–24 July, 1988. Channeled spectra were observed over a bandwidth of 420–690 nm, and the autocorrelation was accumulated for 16 seconds in each case (a total of 4094 frames) before being Fourier transformed. No attempt was made to actively track the fringes, but ‘snap-shots’ were taken to locate the delay. He found that it was possible to detect fringes with as little as 2–4 photons per frame.

2.7 Conclusion

The methods of fringe detection and tracking have been reviewed, considering only those that employ active wavefront tilt correction. Fringe tracking allows an interferometer to operate with greater precision and sensitivity. Larger bandwidths are then possible, and fainter objects can be observed. The tracking permits operation at longer baselines as well, where envelope measurements would be difficult because of large path difference changes.

It was shown that with envelope observations the sensitivity of an interferometer is limited because narrow bandwidths must be used. It is therefore desirable to track the delay, and bring it closer to the zero offset. Pathlength modulation was discussed as a method of estimating the phase of the white-light fringe, permitting bandwidths of up to 20 nm to be used for data collection.

Group delay tracking was then presented, and different methods of data processing were discussed. This form of tracking has several advantages over the others: there is no 2π ambiguity in the location of the fringe position, it yields a much larger coherence length for fringe detection, and it will allow operation at extremely low light levels. Whereas pathlength modulation fails for photon fluxes of ~ 30 photons per frame (Shao and Staelin, 1980), group delay observations should allow active fringe tracking at ~ 10 photons per frame and passive detection at 3 photons per frame.

The theory of group delay tracking will be examined in detail in Part III of the thesis, but now let us turn to Part II and the description of the PAPA camera.

Part II

The PAPA Camera

Chapter 3

The Design of PAPA Cameras

3.1 Introduction

A PAPA camera is a fast photon-counting array detector that was developed at Harvard University in the early 1980s for use in speckle interferometry (Papaliolios and Mertz, 1982; Papaliolios et al., 1985). It provides photon by photon imaging with high time resolution, preserves time-of-arrival for each photon event, and will operate with count rates as high as 1 million photons per second if not restricted by data handling. The camera is well suited for work in speckle and Michelson stellar interferometry, where it is necessary to make short exposure images through a turbulent atmosphere.

Although the acronym *PAPA* is said to stand for *Precision Analog Photon Address* camera, it no doubt owes more to the name of one of its inventors, Costas Papaliolios, and to the previous existence of a MAMA camera (Slater et al., 1990). A PAPA camera was first used for astronomical observations in November 1983, and has since been the main detector for the speckle group at the Harvard-Smithsonian Center for Astrophysics (CfA). It has been used by them to observe an optical source near T Tauri (Nisenson et al., 1985), the α Orionis triple system (Karovska et al., 1986), the Halo binary μ Cas (Karovska et al., 1986), NGC 1068 and NGC 4151 (Ebstein et al., 1989), and SN 1987A (Nisenson et al., 1987; Papaliolios et al., 1989; Karovska et al., 1989). While on loan to other groups it has also been used to study the asteroid 4 Vesta (Drummond et al., 1988), and pluto's moon Charon (Beletic et al., 1989).

In 1988 one of the co-developers, Peter Nisenson, arranged to build five cameras as a cooperative effort with other research groups, including the University of Sydney. These were based on the camera design of Adaptive Optics Associates (1988), also described by Gonsiorowski (1986). They incorporate a dual-stage image intensifier, a large collimating lens, numerous small zoom-lenses, and a single large mask plate. Construction began at Harvard in 1989 and was completed in late 1990. However, the

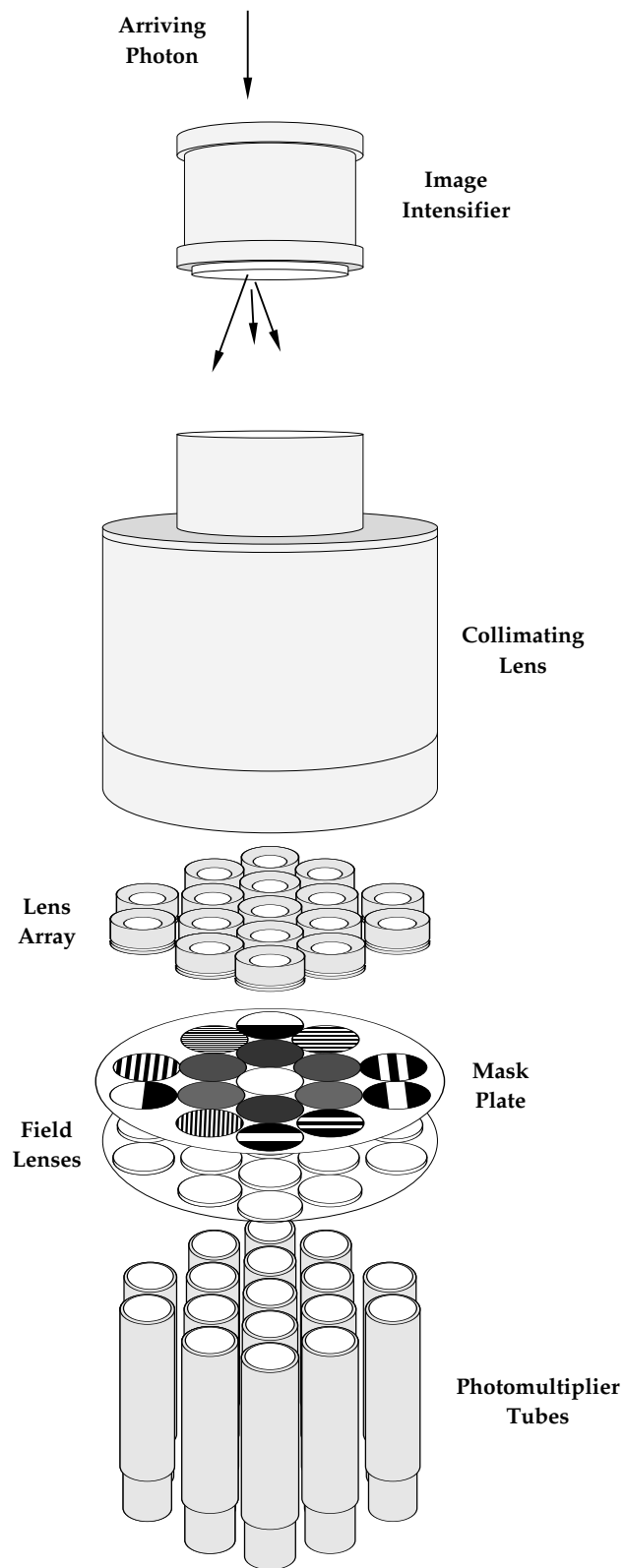


Figure 3.1: The layout of the PAPA camera optical components. Arriving photons are represented by bright photon events at the output of the intensifier. As they arrive, one by one, they are re-imaged onto an array of masks to optically encode their coordinates.

camera received in Sydney proved to have optical and mechanical faults that made it unsuitable for astronomical research. It was therefore dismantled, redesigned, and rebuilt to eliminate vignetting artifacts, to improve the optical alignment, and to include improved analog electronics. The new camera was completed in July of 1992. Its design and performance are presented in Chapt. 4.

This chapter contains a review of the design principles of the PAPA camera. The theory of operation is described in §3.2, including a description of the Gray coded masks and the optical alignment of the system. The use of image tubes, the magnification tolerances of the small lenses, and the choice of their focal lengths are presented in §3.3. The history of the design, from the earliest prototype to the present model, is reviewed in §3.4, and finally, the artifacts in PAPA camera images are detailed in §3.5, where the symptoms and methods of correction are given.

3.2 Address Decoding with Optical Masks

3.2.1 Principle of operation

The front end of the PAPA camera is a cooled single or dual-stage image intensifier. Photons arriving at the input photocathode produce bright spots, or photon events, at the output phosphor. It is assumed that single photon events appear one by one at the output, so that over a short sample time the output of the intensifier is a uniformly dark field containing only one photon event. The camera is designed to determine the location, or address, of the photon event in that aperture.

The output of the intensifier is imaged onto an array of masks. There is an image and mask for each bit of address information, both x and y , and a single clear mask used as a strobe to indicate the presence of a photon event. Therefore, to define a field of 2^m by 2^n pixels, $m + n + 1$ masks are required.

Each mask represents one bit in the address of a pixel. The masks are opaque where the address bit is off, and transparent where it is on. The masks change state as the addresses progress along the x or y axis of the image and appear as sets of horizontal or vertical stripes. Whether or not light from the photon event passes through a mask determines the state of a bit in the event's address: if it passes through then that bit is *on*, and if it doesn't then that bit is *off*. By observing all masks simultaneously the event's address can be determined.

3.2.2 Optical design

Figure 3.1 shows a component view of the optics of a PAPA camera. Although the number and arrangement of masks may differ from one camera to the next, the optical design is essentially the same. The following description, including the dimensions quoted, is typical of most cameras.

Photon events are produced from a 25 mm diameter second generation image intensifier containing a microchannel plate. The high gain of the microchannel plate ensures that each photon event is represented by tens of thousands of photons. The P-47 phosphor is used on the output for its short decay time, improving the speed of the camera and minimizing the chance of event coincidences. The image tube is followed by a high quality, large diameter lens (95 mm or greater), which collimates the light. An array of lenses is then packed within the effective aperture of the large lens, so that each produces its own separate image on the mask plate which follows. The magnification of each image is determined by the ratio of the effective focal lengths of the array lenses to that of the collimating lens, typically 0.5, with the focal lengths of the array lenses carefully matched to meet the required tolerance. The mask plate contains all the masks as separate chrome patterns, deposited by optical micro-lithography. They are designed for an image size of about 10 mm square, and thus an appropriate square or circular field stop is placed at the image intensifier.

The final stage of the optical system is for photon detection. A field lens is placed behind every mask to image the aperture of each array lens onto the surface of a photomultiplier tube, as is used in radiometer design (Driscoll, 1978). The output of each tube is then converted to voltage pulses which are processed by analog and digital circuitry to yield the photon address.

3.2.3 Alignment of Gray coded masks

In the PAPA camera the Gray code is used for the design of the optical masks. This helps reduce the magnitude and frequency of address errors. The Gray code is a binary code that has the remarkable property that only one bit changes in going from one sequential number to the next (Press, 1992). This implies that only one mask edge defines the border between any two pixels. If a photon event lands at a mask edge an address error may occur, but the result of that error is an address that is only one pixel removed from its true address. The uses and properties of Gray codes have been reviewed by Gardner (1972). Heath (1972) also discusses cyclic permuted (Gray) codes and how they were used in Emile Baudot's first telegraph, exhibited at the Universal Exposition in Paris in 1878. The code which Frank Gray re-invented in the 1940's was intended for use in pulse code modulation, and is now commonly found

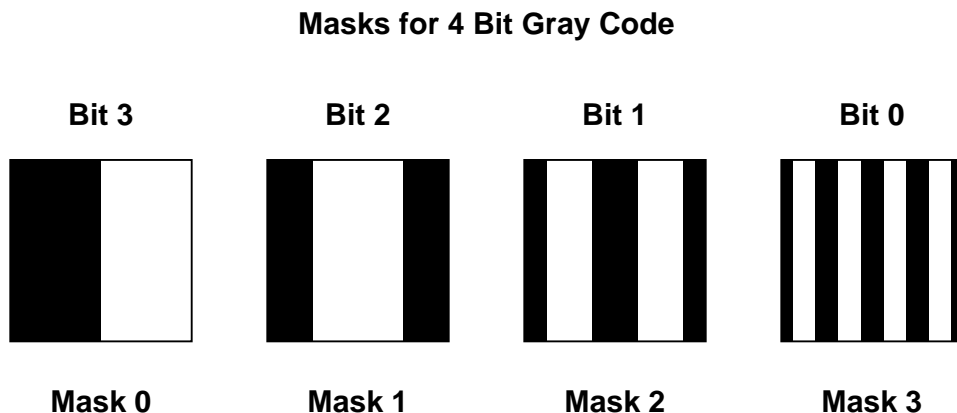


Figure 3.2: Gray coded masks for 4-bit address decoding.

in shaft encoders and similar devices. The Gray code and its relationship to normal binary is illustrated in Figs. 3.2 and 3.3.

The images must be aligned on the masks for the address decoding to be successful. The edges of mask stripes define pixel boundaries, and therefore the image on each mask must be located appropriately to within a small fraction of a pixel. Hence it is crucial for the magnification of each array lens to meet strict tolerances; if they do not, the mask edges will occur at incorrect intervals and alignment will not be possible.

Let us consider an 8-bit camera with 17 masks: 8 masks for each of the x and y axes, and a single clear mask as a reference strobe. The Sydney University camera uses this design, and the layout of its mask plate is illustrated in the next chapter, in Fig. 4.2. The alignment of these masks is presented in Fig. 3.4 where each mask is shown aligned with respect to the finest and coarsest masks. The coarsest mask, which divides the field in two and bisects an opaque stripe of the finest mask, defines the center of the image. All other masks are aligned with reference to the center. The masks are also shown in alignment at the top of Fig. 3.9, but without the pixel boundaries indicated. Note that the stripes of the finest mask are two pixels wide.

The alignment can be observed by illuminating selected pairs of masks and back-projecting their images through the lens system onto the output surface of the image intensifier, or onto a fiber-optic faceplate located there (Papaliolios and Mertz, 1982). The relative alignment of masks can be found in the overlapping images and viewed with a traveling microscope. The alignment is done by moving the mask images with respect to each other according to the alignment chart of Fig. 3.4. The masks belonging to one axis are aligned by rotating the imaging lenses in front of each mask. If the lens's mechanical axis and optical axis do not coincide, as is generally the case

Base 10	Binary	Gray
0	0 0 0 0	0 0 0 0
1	0 0 0 1	0 0 0 1
2	0 0 1 0	0 0 1 1
3	0 0 1 1	0 0 1 0
4	0 1 0 0	0 1 1 0
5	0 1 0 1	0 1 1 1
6	0 1 1 0	0 1 0 1
7	0 1 1 1	0 1 0 0
8	1 0 0 0	1 1 0 0
9	1 0 0 1	1 1 0 1
10	1 0 1 0	1 1 1 1
11	1 0 1 1	1 1 1 0
12	1 1 0 0	1 0 1 0
13	1 1 0 1	1 0 1 1
14	1 1 1 0	1 0 0 1
15	1 1 1 1	1 0 0 0

Figure 3.3: A comparison of binary and Gray code: In going from one Gray coded number to the next only one bit changes state, whereas in normal binary it is common for many bits to change.

for commercially available lenses, rotating the lens will cause a mask image to move in a circle. This provides the range of adjustment needed in either the x or y direction.

Further range could be provided using plane parallel plates of glass placed at an angle between an array lens and its mask. They, however, introduce spherical aberration in the images (Smith, 1990), and are unnecessary if the lenses themselves provide adequate range.

3.3 Design Parameters

There are several design parameters of the camera worth discussing. Of foremost importance is the image intensifier as it is the limiting factor of system sensitivity. The magnification tolerances of the small lenses will then be discussed: it is because of the severe tolerances required of these lenses that so few successful cameras have been built. The proper choice of focal lengths for the collimating and array lenses will then be reviewed. It will be shown that the large field angles necessary for an optically fast design almost preclude the use of achromatic doublets.

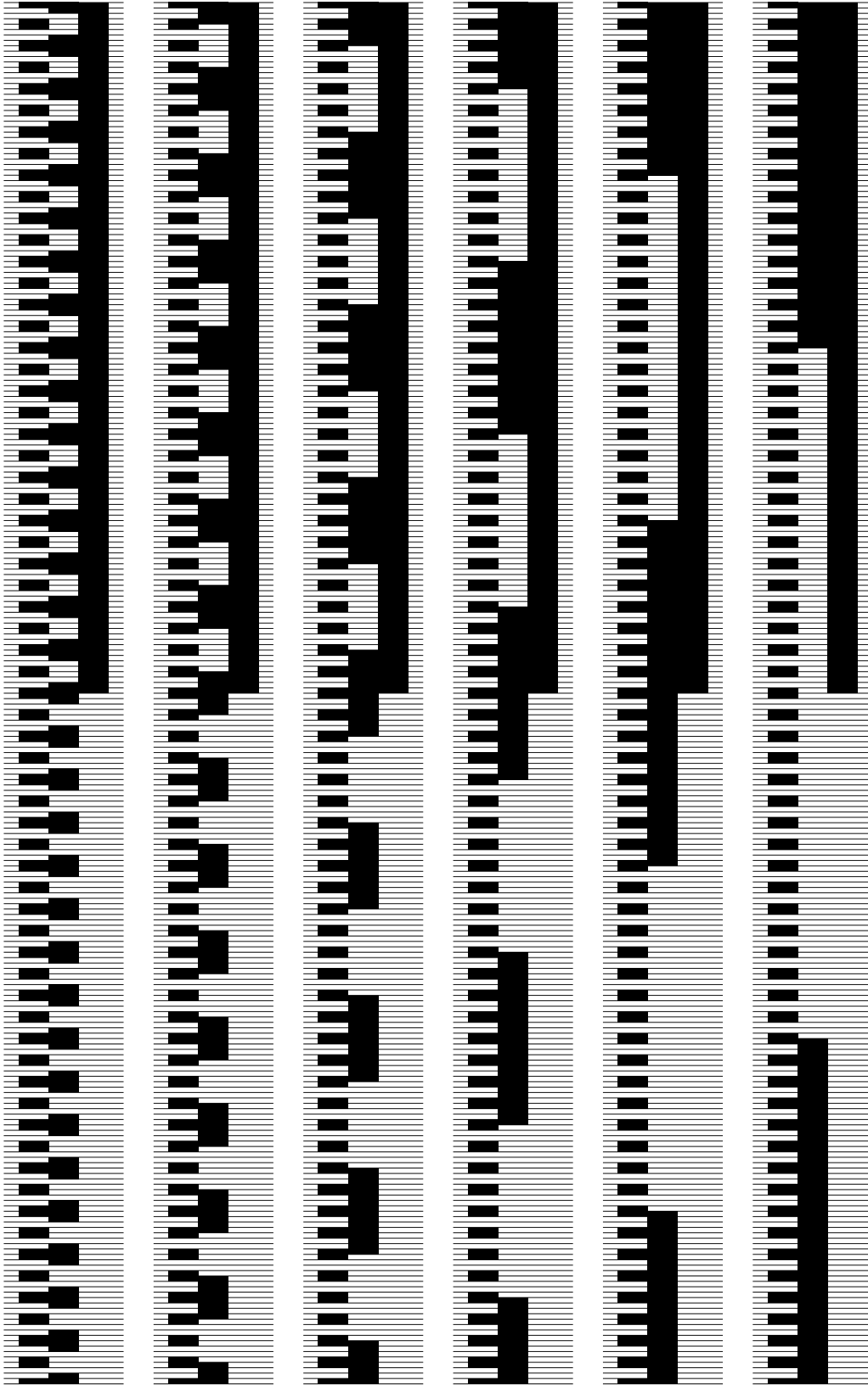


Figure 3.4: The alignment chart. This shows each mask properly aligned with respect to the 0 order and 7th order masks. Every second pixel boundary is defined by the stripe edges of the finest mask, the 7 mask. Half of the remaining mask boundaries, those which lie within the transparent stripes of the 7 mask, are defined by the edges of the 6 mask. All other pixel boundaries lie within the opaque stripes of the 7 mask and are defined by the edges of the 5, 4, 3, 2, 1, or 0 masks.

3.3.1 Image tubes and photon counting

The choice of image intensifier influences the camera's resolution, noise properties, dynamic range, detective quantum efficiency, and the pulse height distribution of photon events. For photon counting applications one would like an image tube with a high quantum efficiency, high gain, and a pulse height distribution which exhibits a strong peak, allowing discrimination against system noise. Both first and second generation image tubes have been used for photon counting.

First generation intensifiers produce photoelectrons from a semi-transparent cathode. These are accelerated across roughly 15 kV, are focused by electrostatic or magnetic fields, and then arrive at a phosphor screen. First generation tubes have a pulse height distribution suitable for photon counting applications, but have too low a gain for them to be useful as a single stage. To achieve high gain the intensifiers are commonly cascaded, with up to four tubes in a chain. Unfortunately, distortion and vignetting may occur, the effective decay time of the phosphor may be lengthened, and image blooming can appear in the region of bright sources. The development of microchannel plate (MCP) technology was seen as a way to circumvent these problems.

The microchannel plate consists of thousands, or millions, of minute glass tubes all packed together in a thin wafer. A microchannel is typically 10 μm in diameter and is coated to act as a continuous electron multiplier. Photoelectrons hitting the channel walls cause an avalanche of secondary events which cascade down the length of the channel. Because of their small size and closed packed structure, the spatial information in the image is maintained. Their characteristics have been described by several authors. Perhaps the best introduction to microchannel plates is given by Corbett (1992) contained in the *Photonics Handbook*. Overviews of the history and development of MCPs are given by Leskovar (1977) and Lampton (1981). The review article by Wiza (1979) contains the results of numerous experiments on MCPs of different architectures. Night vision systems with MCP intensifiers are discussed in the review by Pollehn (1980). The book *Image Tubes* by Illes Csorba (1985) is also a useful reference. Current developments in MCP technology are covered yearly in the conference proceedings of the Society of Photo-optical and Instrumentation Engineers (SPIE); the proceedings of the 1989 conference held in Los Angeles, Vol. 1072, and the 1990 conference in Santa Clara, Vol. 1243, are particularly interesting.

Image intensifiers employing microchannel plates are termed second generation. With a microchannel plate high gain can be achieved in a single stage, obviating the need to cascade the tubes. Furthermore, the appearance of veiling glare, due to electrons scattered from the phosphor, is largely suppressed. Gains even as high as 10^8 are possible in some configurations. Not surprisingly, modern photon counting detectors make wide use of microchannel plate technology, whether they be incorporated into

intensified CCD systems, attached to resistive anode arrays, or as part of the PAPA camera.

Pulse Height Distribution

The production of secondary electrons is governed by statistical phenomena, and consequently the brightness of photon events produced by the image tube may change from one event to the next. For photon counting applications one would like to have all photon events of the same intensity, and for that reason it is desirable to use a microchannel plate under high gain.

When a microchannel plate is biased for low gain its pulse height distribution resembles a negative exponential, but under high gain the space charge in the microchannels becomes saturated and the pulse height distribution exhibits a peak with a well defined width (Slater, 1990; Leskovar, 1977). Most pulses are represented by the same number of accelerated electrons. This permits the discrimination against noise events, a useful feature for photon counting. Space charge saturation occurs when there are so many electrons already in a channel that they drive other newly released electrons back into the channel walls, limiting the available gain. Unfortunately, in straight channel MCPs ion feedback will set in prior to space charge saturation.

At very high bias voltages, if the cloud of secondary electrons becomes dense enough, the residual gas molecules in the channels will be ionized, including some molecules from the *walls* of the channels. The positive ions are accelerated back towards the input of the plate, giving rise to after-pulsing. The feedback occurs on timescales much shorter than the decay time of the phosphor, and may result in unusually large pulses. If the feedback is unchecked then the image may become very noisy and the microchannels may eventually be destroyed. The push towards higher gain has therefore led to the development of microchannel plates with different architectures: these include curved, chevron, and z-stacked configurations (Wiza, 1979; Slater et al., 1990). The absence of a straight channel in these cases acts to limit the mean free path of the ions and therefore suppresses the feedback. Curved channel MCPs permit operation in space charge saturation mode, and are used specifically for photon counting applications (Slater et al, 1990).

Detective Quantum Efficiency

When discussing image intensifiers the term Quantum Efficiency (QE) is used to describe photocathode sensitivity (Csorba, 1985). The performance of the different types of photocathodes have been reviewed by Pollehn (1980) and Csorba (1985). The multialkalide S-20 photocathode, used in all first and second generation image tubes, will have a peak quantum efficiency of between 10% and 15% at 550 nm. Third

generation image tubes are distinguished by the different photocathode they use, a multi-layer gallium arsenide (GaAs) compound. They are capable of much greater quantum efficiencies, with peak QE of 25%, extending from 650 nm to 850 nm.

However, it is the Detective Quantum Efficiency (DQE) of the intensifier that describes the system sensitivity. It accounts for losses within the image tube, as may be produced by the presence of a microchannel plate. The MCP introduces a filling factor, quantified by an *open-area ratio*. A large fraction of the electrons that are accelerated from the photocathode impact on filled areas of the MCP. Geometrical constraints allow only about 57% of the photoelectrons to land directly within the microchannels. The remainder either never initiate secondaries, or are reflected to land elsewhere, perhaps some distance from the original impact point (Pollehn, 1980). The image transfer characteristics and the Detective Quantum Efficiency are therefore reduced. Moreover, MCPs are frequently coated with a protective film to prevent damage to the cathode caused by ion feedback, and this further reduces the sensitivity. Attempts to improve the DQE of these devices have included flaring the inputs of the microchannels to increase their open area ratio, applying a bias voltage to draw electrons into the channels (Lampton, 1981), and removing the thin film on the input of the MCP (Airy et al, 1990). Although unfilmed MCPs have better sensitivity, they have a reduced life expectancy.

Image tubes and PAPA cameras

The microchannel plate is needed for its high gain and fast response time, but its use almost halves the available quantum efficiency. With PAPA cameras only one approach has been used to overcome this problem.

The strategy, first discussed by Papaliolios and Mertz (1982), is to employ a hybrid combination of first and second generation intensifiers (also termed Gen I and Gen II). It was reasoned that although the Gen I has only modest gain it would be enough to drive the Gen II and overcome losses at the microchannel plate. The result would be a system with a DQE approaching that of the first stage, but with the gain of the second. In this format it would also be possible to discriminate against noise originating in the Gen II alone, since events from the Gen I would be represented by much larger pulses. This approach has also been used by Foy and Blazit in the photon-counting CP40 detector (Foy, 1988), an intensified CCD camera.

It is obvious from the previous discussion that the hybrid approach is very different from designs which attempt to optimize single stage MCP detectors for photon counting. The disadvantages of a hybrid system are common to other cascaded arrangements: much larger voltages are required across the image tube assembly, and the intense electric fields around each stage may cause coronal discharge if the tubes are not correctly coupled; veiling glare, which is larger in first generation intensifiers

(Pollehn, 1980), will be amplified by the second stage; and geometric distortion is also possible, a fact acknowledged by Papaliolios et al. (1985). Attempts to improve the DQE based on hybrid systems have met with mixed results. Nevertheless, Latham (1982) reports that up to 80% of the photocathode sensitivity can be obtained. He describes a similar strategy for use with photon-counting Reticons and outlines the difficulties that he encountered. In the Sydney University camera, to be described later, only a single Gen II intensifier is used.

3.3.2 Magnification tolerances

Another limiting factor in the design is the proper choice of array lenses. Their effective focal lengths have tolerances determined by the magnification tolerances between mask images. Let us consider an array lens and collimating lens with effective focal lengths f_a and f_c respectively. When light is propagated forward in the system, from intensifier to mask plate, an intensifier of output field diameter d will subtend an angle $\theta = d/f_c$ as viewed through the collimating lens. Therefore, after re-imaging with the array lens, the diameter of the field on the mask plate will be $f_a\theta$, or $d(f_a/f_c)$. The magnification of each field is therefore determined by the ratio of focal lengths of the array lens to that of the collimating lens. Since the same collimating lens is used by all array lenses, measured percentage differences in magnification from one image to another are due to the same percentage differences in the effective focal lengths of the array lenses.

All the mask edges, when back-projected to the image intensifier should be located appropriately to within a small fraction of a pixel. If n is the number of masks along one axis, d is the aperture size at the image tube, then the pixel size will be $d/2^n$, and the allowable position error of mask edges at the image tube will be $\beta d/2^n$, where β is the (unitless) fraction of a pixel tolerated.

Now, if ξ is a coordinate at the mask plane which is normal to the direction of the mask edges, M is the correct magnification from mask to image tube, and ϵ is an error in magnification, then the following must be true across the whole image for a proper alignment:

$$\xi(M + \epsilon) - \xi M < \beta \frac{d}{2^n}. \quad (3.1)$$

If the center of the field is perfectly matched then the worst-case errors will occur at the edge of the field, where $\xi M = d/2$, or equivalently $\xi = d/(2M)$. Putting this value for ξ in Eq. 3.1 we can now write:

$$\frac{\epsilon}{M} < \frac{\beta}{2^{n-1}}. \quad (3.2)$$

It follows that if 8 masks are used for 256 pixels and $\beta = 0.2$, then the focal lengths of all the array lenses must be the same to within 0.15%. By contrast the focal lengths of commercially supplied achromats are usually quoted with tolerances of only 1% or 2%.

3.3.3 The field size of images

Achromatic doublets are almost precluded from use as array lenses, since they only image well with a field radius of less than 2.5 degrees. If the output of the intensifier has a diameter of 25 mm then the collimating lens must have a focal length greater than 300 mm for the achromats to be useful. Even then they must be used with field flatteners. Although there exist collimating lenses which would allow the use of achromats, the magnification tolerances would make the achromats expensive.

If θ_{\max} is the maximum field of view of the array lenses, d is the diameter of the field at the intensifier, and f_c is the focal length of the collimating lens then we must have

$$f_c \geq \frac{d}{\theta_{\max}}. \quad (3.3)$$

Only the camera built by Papaliolios et al. (1985) used achromats and field flatteners. For their design they used a collimating lens with an effective focal length of 285 mm. All other cameras have used the miniature zoom lenses made by Rolyn Optics and collimating lenses with shorter focal lengths.

3.4 Historical Background

3.4.1 Papaliolios and Mertz (1982)

The earliest version of the camera (Papaliolios and Mertz, 1982) operated with 256 x 256 pixels, or 8-bits for both x and y axes. It used two 135 mm f/2.8 telephoto lenses (joined front to front) to image through an array of prisms, onto masks that were attached to the prism surfaces. The telephoto lenses would have produced a single 1:1 image had not the prisms been there; the 18 prisms each acquired a pie-slice of the converging beam, and used that light to form 18 separate images—only 17 of the 18 prisms were used. However the prisms vignetted the converging beam and produced images that made it impossible to correctly flat-field the camera. This design was therefore abandoned. Nevertheless, some of the basic components, including the Hamamatsu type R647-04 photomultipliers and the Varo model 3603 intensifier, have been used in all other cameras.

3.4.2 Papaliolios, Nisenson, and Ebstein (1985)

A subsequent design was published several years later (Papaliolios, Nisenson, and Ebstein, 1985) this time to be operated with 9-bit resolution, 512 x 512, with the option of averaging to 8-bits. It performed the light division in a collimated beam, rather than a converging beam. A 4 inch diameter f/2.8 Kodak Aero-Ektar lens was

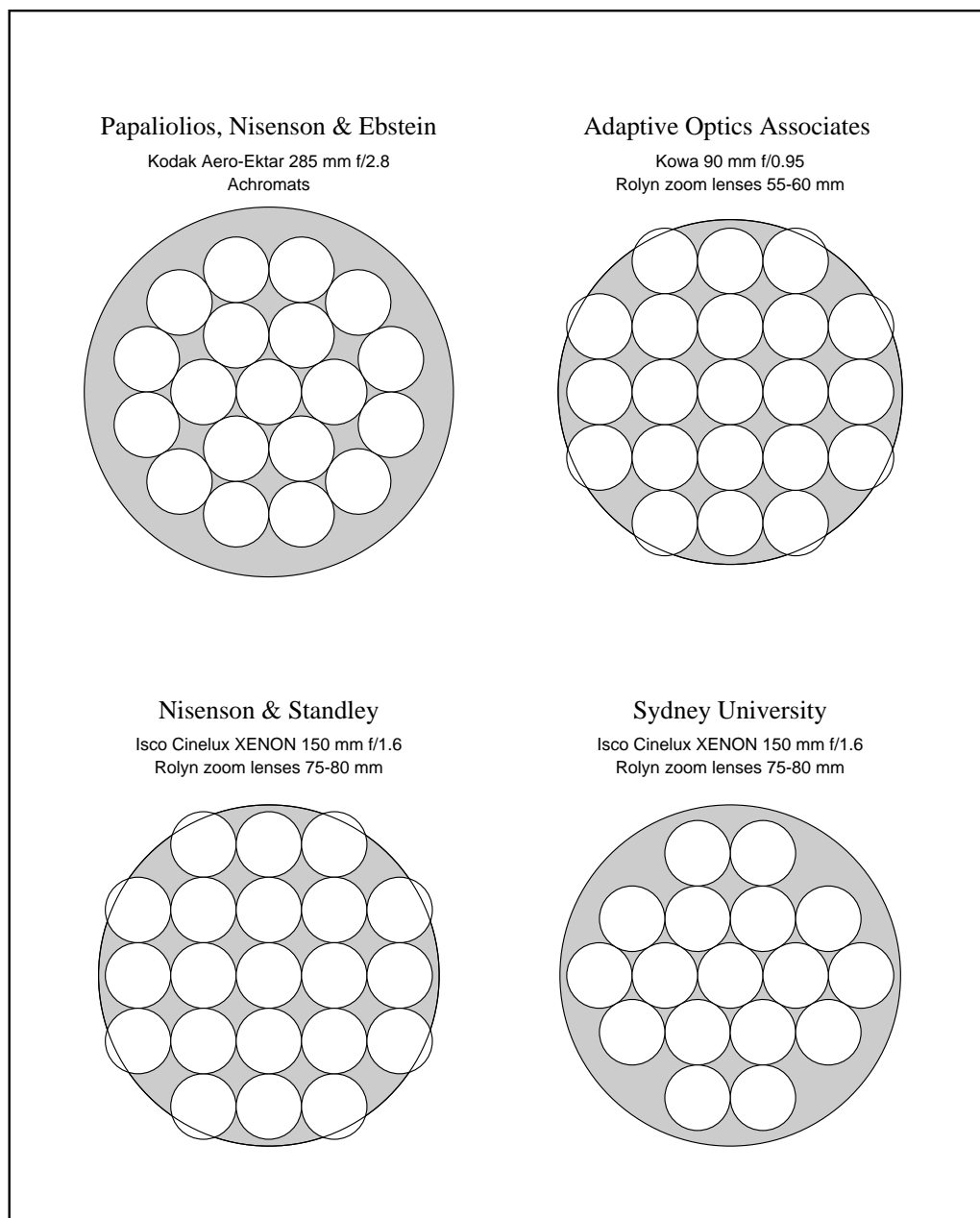


Figure 3.5: The lens arrays for the PAPA cameras: The large circle in each picture bounds the effective aperture of the relevant collimating lens. In each case the effective aperture is determined from the focal length and f number. The original camera did not use lens arrays, and is therefore not included in this figure.

used to collimate the light and to image through an array of achromatic doublets, with field-flatteners located just in front of the mask. As in the previous case the mask for each bit was made separately and aligned by physically moving it into the correct place with respect to other masks. The Aero-Ektar lens has a focal length of 284.5 mm (11.2 inches), which is implied by its diameter and f number. The achromats were arranged in two concentric circles about a central lens; the inner circle having 6 lenses, and the outer circle having 12 lenses, for 19 lenses in total. This packing is shown in Fig. 3.5. The Aero-Ektar lenses were large aperture lenses used during World War II as part of aerial surveillance cameras. After the war they were easily obtainable, but they are now no longer manufactured.

Although this camera was successful, as illustrated by the images in the 1985 *Applied Optics* paper, it proved extremely difficult to align, since all masks had to be individually positioned for rotation as well as either x or y movement. Only one such camera was made. It is now owned by Richard Goody and Costas Papaliolios of Harvard, and is currently operated by James Beletic for Georgia Tech Research, in Atlanta.

3.4.3 Adaptive Optics Associates (1988)

Following the success of the 1985 camera the firm of Adaptive Optics Associates (AOA) of Cambridge Massachusetts was contracted by Harvard to produce three more. AOA redesigned the optics to make it easier to align, and it is this design which has been the model for all subsequent cameras. It has been described by its user's manual (Adaptive Optics Associates, 1988) and in the paper by Gonsiorowski (1986). They used a Kowa CL79 90 mm $f/0.95$ *Ultra High Speed Lens* to collimate the light from the intensifier, and used 21 small zoom lenses packed in a square array with a central 9 in a 3×3 pattern, flanked on each of four sides with 3 lenses. This packing is shown in Fig. 3.5. The effective aperture diameter for the collimating lens was slightly less than 95 mm. They included some important changes from the previous design.

- Adjustable focal length zoom-lenses, produced by Rolyn Optics, were used as the array lenses, rather than achromats. This made it possible to set the focal length of the lenses to the tolerance discussed in §3.3.2. These lenses had focal lengths adjustable in the range of 55–60 mm.
- Microlithography was used to produce all the masks on a single plate of glass. So that for an alignment of the camera, rather than moving a mask with respect to the image, as was done for previous alignments, the image was moved across a mask; tilted glass plates in the converging beam of the array lenses were used to accomplish this, as described in Appendix C.
- A combination of *two* image tubes were used to attempt to boost the Detective

Quantum Efficiency (DQE) of the system from 2% to near 10% or possibly 15%, for reasons described in §3.3.1.

Although a 10th mask was included in each dimension—of the same scale and in quadrature with the 9th mask—only 9-bits of data were produced for each axis, and in practice these were averaged to 8-bits. There were, however, 21 masks in total.

One of the cameras is held by Peter Nisenson at the Harvard-Smithsonian Center for Astrophysics, while another is used in the Mark III interferometer at Mount Wilson, as part of their angle tracking system as described by Clark et al. (1986).

3.4.4 Standley & Nisenson (1989)

In 1988 Harvard (Standley & Nisenson, 1989) undertook to produce five PAPA cameras as a cooperative effort with four other institutions: Imperial College of London, Georgia State University, Jet Propulsion Labs, and the University of Sydney. The attempt was partly to reduce the per-unit cost, and partly to improve the shortcomings of previous efforts. The cameras were essentially the same as those produced by Adaptive Optics Associates, with a few minor changes: A 150 mm f/1.6 Isco Cinelux-Xenon lens was used as the collimating lens, and different Rolyn zoom lenses were used, this time having adjustable focal lengths of 75–80 mm. The same number of lenses were used in exactly the same square array as before.

However, there were problems with this design. The Astronomy Department at Sydney University purchased one of these cameras, and it became clear that it would not be a useful detector without modifications to its optics. The optical problems are illustrated in the flat field of Fig. 3.6, and in the integrated cross sections of Figs. 3.7 and 3.8. This image represents a typical flat field response of the camera. There are obvious misalignment and vignetting artifacts in the image. It has a tartan pattern across it caused by defocusing of the array lenses and the shifting of their positions. Along the top and to the right are fragments of mask images, present because the outer lenses in the array were vignetted by the collimating lens. The light was also vignetted *after* the masks, before detection by the photomultiplier tubes: the field lenses had been recessed too far into the photomultiplier housing, on a support which blocked light from the outer edges of the field. The image is of limited extent because of a vignetting stop at each field lens. The vignetting on the masks meant that the camera could never be flat-fielded. The misalignment meant that the camera would not be useful even if the vignetting were corrected.

The main problems were associated with the construction of the lens-mask assembly. Some zoom lenses were not of sufficient quality to produce distortion-free images. Most had not been adjusted to within the required tolerance, making alignment artifacts inevitable. The collimating lens did not have a large enough effective diameter

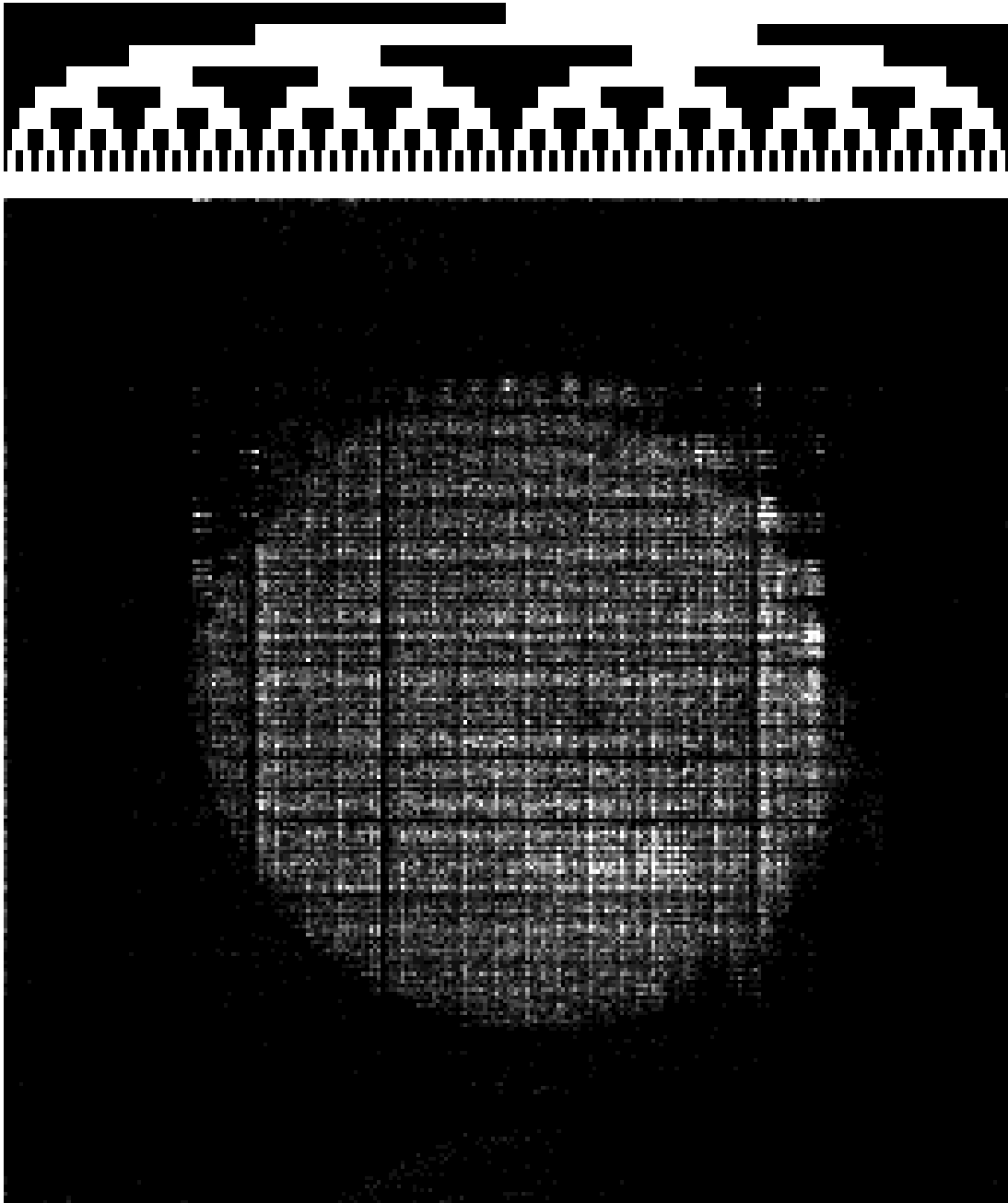


Figure 3.6: Flat field from the camera received at Sydney University from Harvard University. There are numerous artifacts in this image, the most visible being the vignetting from the field lenses. The image should span the full range of addresses, but has been truncated to a smaller circle because the field lenses were too far away from the mask plate. Vignetting by the collimating lens is also apparent in the corrugation features on the top right of the field, showing a folding of photon data across the X1 mask edge. The dark vertical lines are due to misalignment of the X1 and X2 masks.

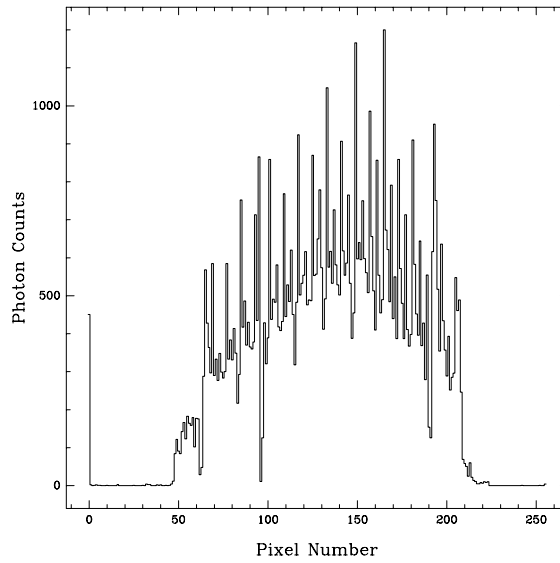


Figure 3.7: An integrated cross section for the x axis of the previous flat field. All the photons that were in the image are represented here. The highest and second-highest peaks are all due to the X5 mask. The X6 mask also causes much smaller peaks to be present. The lowest cut into the image on this axis is due to one edge of the X2 mask, at pixel 96. The X1 mask also cuts two edges, one near pixel 64 and the other at 192.

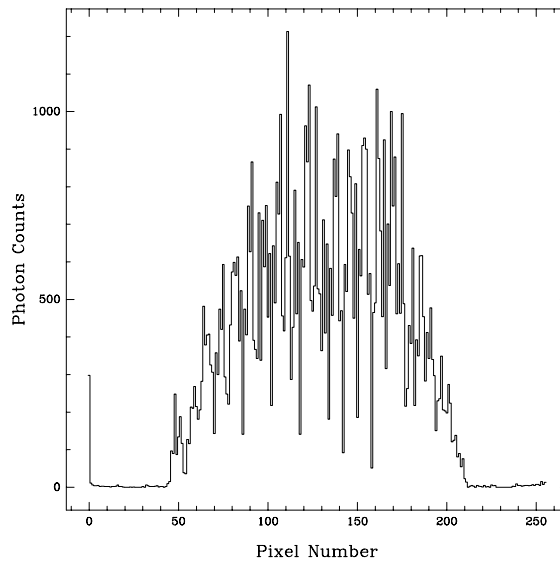


Figure 3.8: An integrated cross section for the y axis of a flat field from the same camera. Defocusing is present in this axis, and almost all masks contribute misalignment artifacts. Most noticeable amongst these are the lower regularly spaced cuts at the edges of the Y4 mask. The lowest is from the Y2 mask at pixel 160, and the second lowest from the Y3 mask at pixel 144.

for the packing of the zoom-lenses, causing the masks to be vignetted. Furthermore, the analog electronics did not have the correct electrical impedance matching, nor shielding to eliminate noise problems. Therefore neither refocusing the image intensifier nor adjustments to the discriminator settings of the electronics, would produce improvements on the features seen in the flat field.

It was decided therefore to rebuild the camera using a different design. This work is discussed in Chapt. 4. The reasons for the redesign are best appreciated through an understanding of the image artifacts. These artifacts will therefore be reviewed in detail.

3.5 Artifacts in PAPA Camera Images

The design of the PAPA camera is deceptively simple, but few working models have been built. The design has been modified several times since the first prototype was described by Papaliolios and Mertz (Papaliolios and Mertz, 1982). Each change has been occasioned by optical and mechanical problems encountered in previous efforts; vignetting, magnification tolerances, and difficulty in alignment are all problems that have had to be surmounted. The success of these efforts is apparent from their results: although cameras have been designed for 1024 x 1024 pixels (Gonsiorowski, 1986; Standley and Nisenson, 1989), and it has been claimed that 4000 x 4000 resolution is possible (Papaliolios et al., 1985), no astronomical observations have been reported in the literature that use a PAPA camera with greater resolution than 256 x 256 pixels. When problems occur with a camera they manifest themselves in image artifacts.

The masks are responsible for the characteristics of the artifacts. The artifacts appear as lines along mask boundaries or as images of the masks themselves. As errors can occur in both axes, they will often cause tartan patterns to appear. The artifacts arise from optical or electronic faults, or from a combination of both. They may be caused by alignment errors, defocusing of lenses, incorrect thresholding in the electronics, vignetting of masks, and faults relating to the performance of the image tube. This discussion will be illustrated with PAPA camera images from the rebuilt Sydney University camera.

3.5.1 Alignment errors

Each mask defines a set of pixel boundaries. In Fig. 3.4 we see that the 7 mask defines 128 pixel boundaries, the 6 mask defines 64 boundaries, the 5 mask defines 32 boundaries, and so on. If one mask is out of alignment then a given set of pixel boundaries will be shifted out of place.

A mask edge determines the relative size of two adjacent pixels at the image

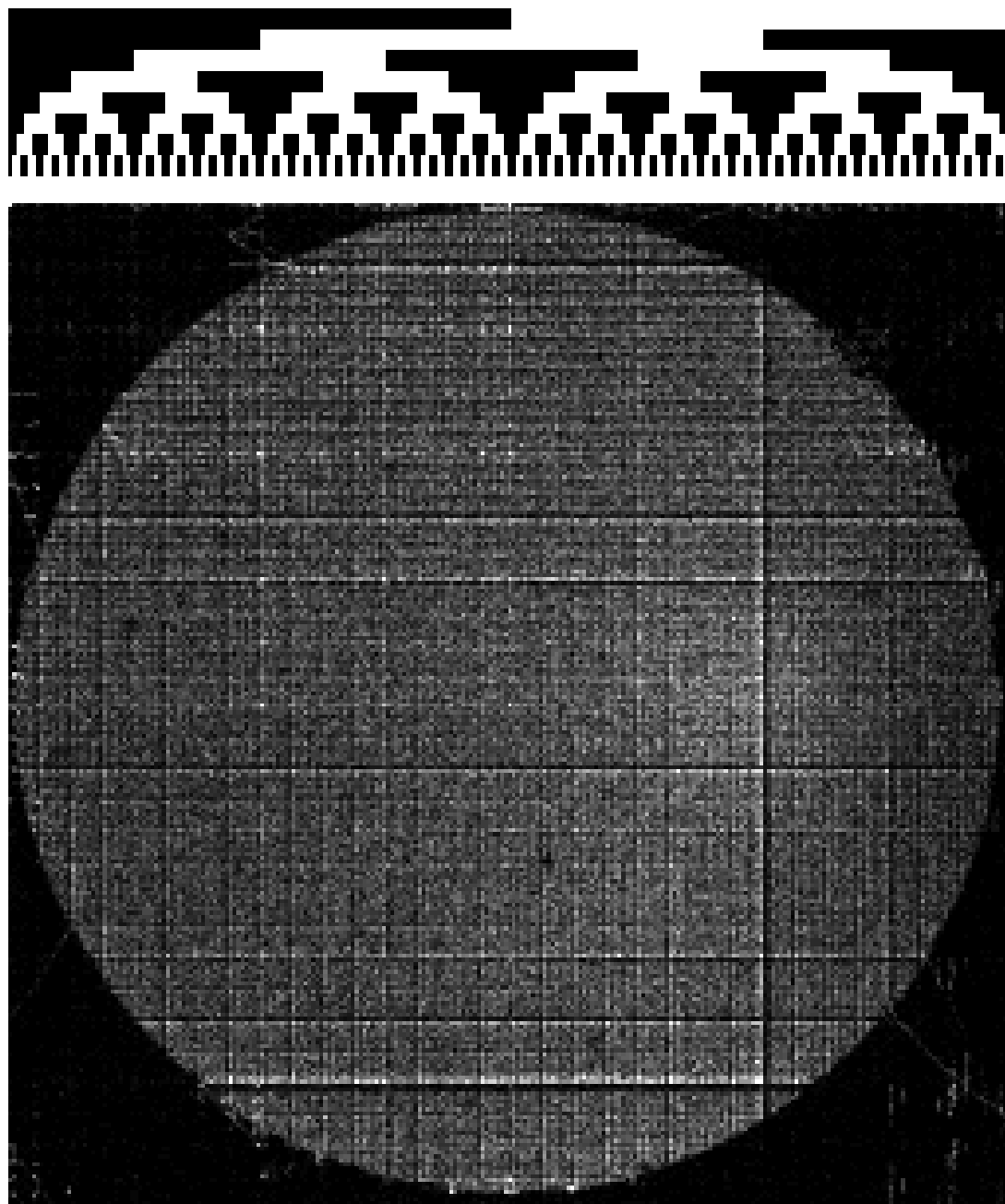


Figure 3.9: Flat field with alignment errors. A circular flat field is shown which contains alignment faults in both axes. The main faults are associated with shifts and magnification errors in the X1, Y2, and Y3 masks. A magnification error in the X1 mask causes only one of two mask edges to be misaligned. The complete Gray code is shown for reference immediately above this image.

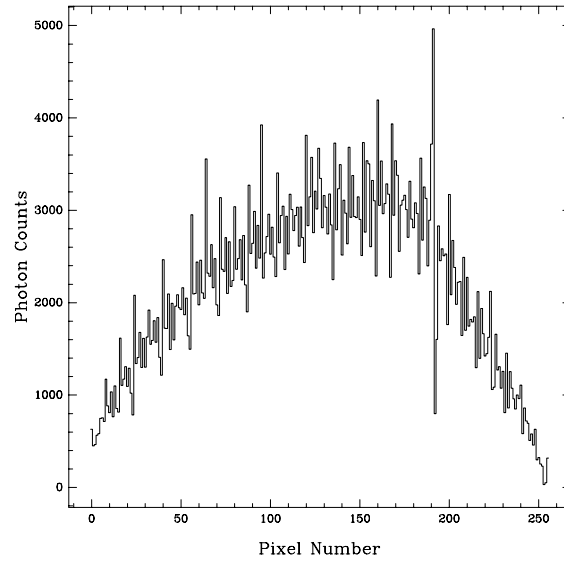


Figure 3.10: A cross section through the x axis of a flat field. The flat field is not truly flat, but has a circular outline that accounts for the main feature of the cross section—the large bulge. The pixel to pixel variations are roughly $\pm 10\%$ of the mean signal level. The most prominent artifact is the misalignment of one edge of the X1 mask.

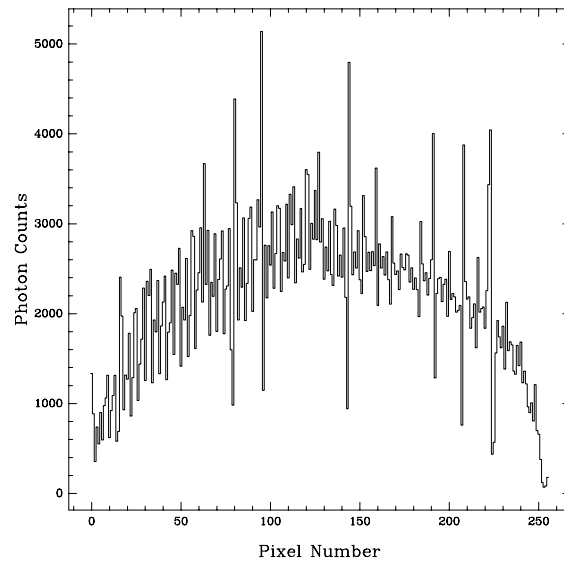


Figure 3.11: A cross section through the y axis of a flat field. This is similar to the previous figure, except that there are more numerous alignment errors, and possibly a defocus error towards lower values of y . The pixel-to-pixel variations are larger than those on the x axis.

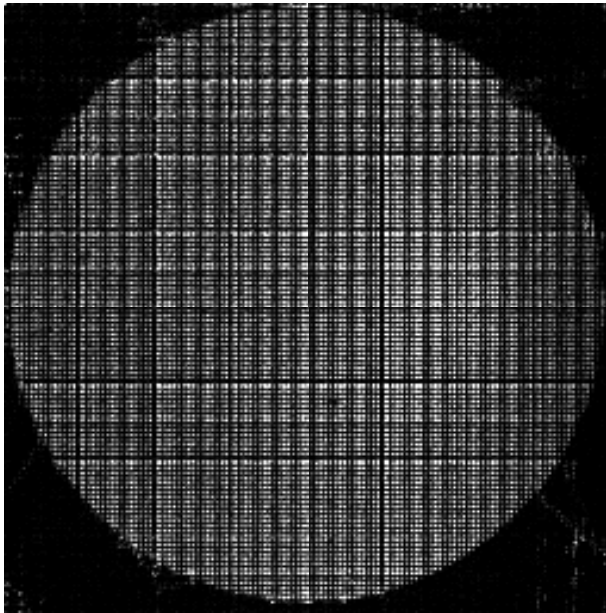


Figure 3.12: Flat field with defocus artifacts. The image tube was moved from the best focus location, and alignment errors are now present at almost every mask edge, producing a very granulated image.

intensifier. If a mask edge is shifted to the right then the pixel to the left will become larger, and the pixel on the right will become smaller; photons that would have been detected by one pixel are erroneously detected by the other. If the shift becomes as large as one pixel's width then one pixel will receive all its neighbor's photon events. This will occur at every pixel boundary defined by that mask.

Alignment errors therefore have a characteristic appearance in the image: they highlight the edges of the mask that has been misaligned. The pixels on one side of an edge will be brighter than those on the other, and both sides will have a response different from the background of the image. The width of the bright and dark lines, measured in pixels, indicates the magnitude of the alignment error. The sequence of bright-dark or dark-bright indicates the direction of the error: a bright line followed by a dark line indicates that the mask is misaligned towards higher addresses. This is illustrated in Figs. 3.9, 3.10 and 3.11, from a flat field which contains magnification errors. If there are magnification differences between array lenses then alignment errors will be present, but the direction of misalignment will change across the image: some mask edges will appear shifted to lower addresses, others will appear shifted to higher ones.

Alignment errors may arise not only because of mistakes in the alignment, but also if either the array lenses or the collimating lens are poorly focused: a defocus will blur the boundary of masks, making the alignment more sensitive to the discriminator

settings (to be discussed); a defocus of the collimating lens will introduce spherical aberration, and each array lens will then see the image tube from a slightly different angle, causing the alignment to change. The image in Fig. 3.12 is a flat field for a defocused image tube. Alignment errors are present at almost every mask edge.

The masks associated with the alignment errors can be identified by placing Gray coded masks, scaled to the full image size, along each axis of the image. This is illustrated in Fig. 3.9 for the x axis. The image artifacts will occur at the mask edges, and those edges can easily be seen against a flat field. After the masks have been identified then the alignment of those channels can be corrected.

Minor alignment problems can also be overcome by rounding address bits (Ebstein, 1987). This technique requires $m+1$ and $n+1$ masks on each axis for a 2^m by 2^n camera, with the extra Gray masks used for the averaging. Normally the finest Gray mask defines only every second pixel boundary. However, an extra mask, representing the next bit of Gray code, could be used to define *every* pixel boundary if the address is averaged in the correct way.

Begin with $m+1$ Gray bits of information for addresses along the x axis. Now convert the addresses from Gray code to binary, add 1 to them, and then throw away the least significant bit. You are left with m bits of address information, but now the pixel boundaries are re-defined so that they coincide with the edges of the extra Gray mask. Because those edges are etched on a single mask, the alignment should be perfect.

3.5.2 Discriminator setting errors

The PAPA camera electronics receive charge pulses from the photomultiplier tubes that are converted to voltage pulses and amplified using fast low-noise op-amps. The pulse height statistics of photon events are determined by the characteristics of the image tubes.

The method of variable threshold discrimination has been applied in PAPA cameras for address decoding in the presence of unfavorable pulse height distributions (Gonsiorowski, 1984). The voltage pulses are passed to comparators whose reference is half the height of the pulse from the clear channel, the strobe. If a voltage pulse is large enough compared to the strobe pulse, then the level on the output of the comparator is set high. The decision level at a comparator is set through the adjustment of a variable resistor.

If a threshold is set too high then addresses will be biased to regions where that address bit is off, and if it is set too low then the bias is to regions where that address bit is on. In either case an image of that mask, either positive or negative, will appear in the data. It can be seen in Figs. 3.13 and 3.14 that the discriminator

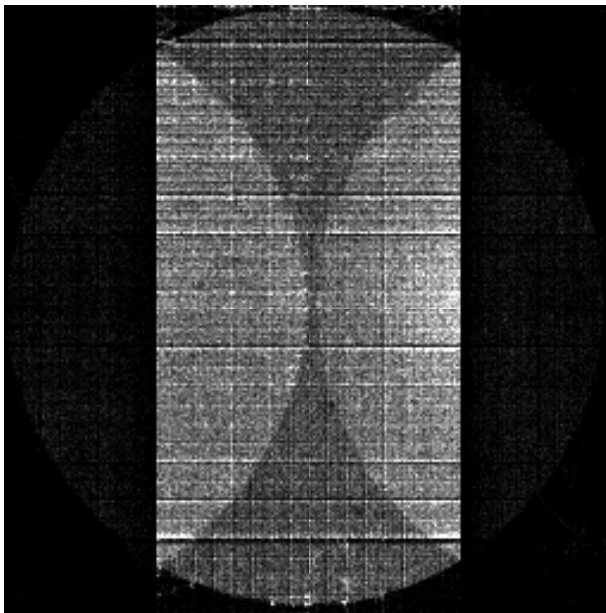


Figure 3.13: Flat-field with a discriminator error in the X1 channel. Because the threshold in the X1 channel is set far too low, 95% of all photon events are given addresses where the X1 bit is set on, and the mask image and folding are clearly seen. For a small error in the threshold only a faint image of the mask would be present.

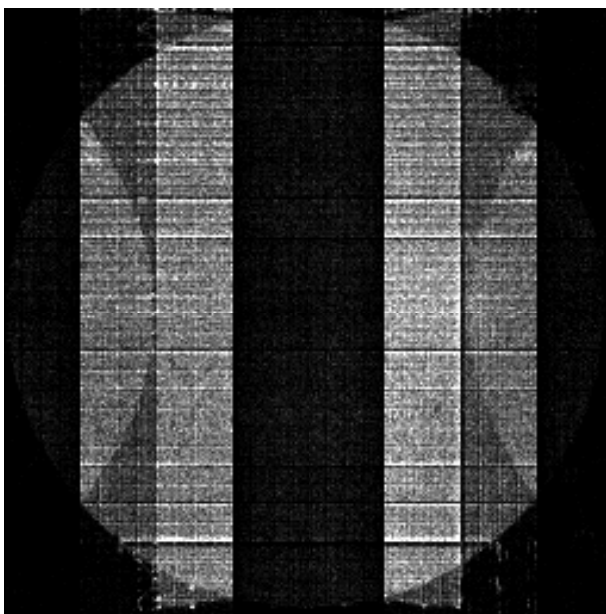


Figure 3.14: Flat-field with a discriminator error in the X2 channel. This is similar to the previous example, but the folding of the data is now more complicated.

errors not only produce a mask image, but also fold the original image into it, cast back across the mask boundaries. The folding is a more subtle aspect of this artifact, and is best understood with an example.

Note that there is symmetry of reflection in the Gray code about the center of the image (see Fig. 3.9). All address bits but the most significant bit have the same state at equal distances from the center. The folding is a result of this symmetry. It implies that for an 8-bit detector if the address bit corresponding to the 0 mask is never on, then all photons will be given addresses less than 128, and the image will appear folded from higher address, back across the edge of the 0 mask. The folding for other masks is more complicated.

The nature of the folding can be seen in the fractal-like geometry of the Gray code. The state of all address bits have a symmetry of reflection about a given mask edge. This is true for a distance of up to half a stripe width either side of that edge. A systematic bit error therefore folds the image across the mask edges. The folds take the data out of one bit-state of that mask and into the other.

In the absence of other errors the camera can be flat-fielded by adjusting the discriminator settings so that the mask images and the folding disappear. This is best done by observing the frequency that each address bit is on. If the camera were to observe a flat field that spans all addresses, then each address bit should be on in 50% of recorded photon events. This will be the case when the discriminators are adjusted correctly. If the flat field is circular, rather than square, then an integration across the mask is necessary to determine the proper thresholds. These errors involve only a small electrical adjustment and are the easiest of all artifacts to correct.

3.5.3 Vignetting errors

Adjustments of discriminator settings have a global effect; they treat all photon events equally, independent of their addresses. Vignetting causes an address dependent error, and when it is present it is not possible to flat-field the camera.

Each of the array lenses should produce an exact replica of the output of the image intensifier. If there is vignetting somewhere in the optical chain then images on some masks will have a variable illumination across them. A photon event that occurs in a region that is vignetted may not be detected by the photomultipliers even if it occurred in a clear region of the mask. Vignetting in one channel therefore biases that address bit off, but only in the parts of the field that are vignetted. This causes an image of the mask to appear there, complete with image folding, as if there were a field-dependent discriminator error. Surprisingly, this may occur even when the bit statistics indicate that the discriminators are set correctly. Changing the discriminator threshold will make different parts of the mask come into view superimposed on the

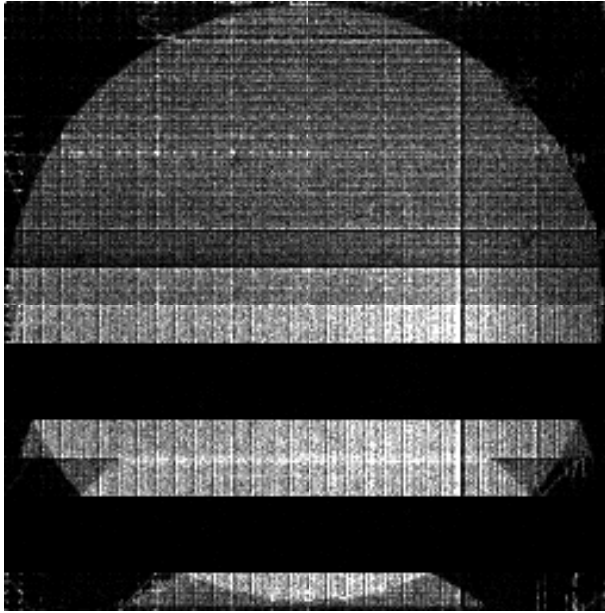


Figure 3.15: Flat-field with a vignetting error in one mask. The discriminator settings are adjusted correctly, but a card was placed to vignette the lower half of the Y3 mask. It now appears as if there is a discriminator error in that part of the image.

image, but will not alleviate the problem. Vignetting is illustrated in Fig. 3.15 where the Y3 mask was partially shadowed.

The presence of vignetting would indicate a serious oversight in the optical design. If vignetting errors are present, and their cause is known, then the optical design of the camera must be suitably modified.

3.5.4 The sum of all masks

If all the aligned masks are back-projected at the same time then the image in Fig. 3.16 should be obtained. This is not a flat field, but a complicated tartan pattern. Dark areas occur in the tartan where most address bits are off (for both x and y addresses), and light areas occur where most address bits are on. This may appear, in whole or in part as a positive or negative image, due to the behavior of the image intensifier.

If the gain of the image tube is too low then single photon events will produce insufficient light to illuminate all photomultiplier tubes. Addresses will be biased to regions where most address bits are off, and photon events will appear to cluster in the darker regions of the tartan. The negative of Fig. 3.16 will appear in the data.

Even if the intensifier has sufficient gain, the tartan will eventually appear in the images as the photon count rate increases. At high light levels the response of the microchannel plate will change; pulses that represent photon events will have a lower average height and become severely distorted. If the intensifier has automatic brightness control then it will reduce the bias of the plate before damage occurs (Pollehn, 1980). If not, then secondary electrons will be removed from the walls of the microchannels faster than they can be replaced, producing current saturation (Pollehn, 1980; Wiza, 1979). The dead-time required to replenish the electrons, determined by the impedance across the microchannels, can be up to 2 ms long (Laprade, 1990). With the lower pulse height distribution the discriminator settings become systematically high, and the negative of the tartan will appear. A flat field with excess light levels is shown in Fig. 3.17.

Event coincidences may also occur at high count rates. The camera's operation depends on there being only one photon event on the image tube during the sampling interval. If there is more than one event present then a composite address will be created, where all the Gray bits that were on in the separate addresses will be on in the composite address—the two addresses will be ANDed together optically (x and y separately). Photon events will be biased towards regions where most Gray bits are on, the brighter regions in Fig. 3.16. This can occur because of true event coincidences, but also from the presence of hot spots on the intensifier—isolated points which are on permanently.

Under high gain conditions ion feedback may be initiated in the microchannel plate. The feedback process occurs on timescales much shorter than the decay time of the phosphor, yielding pulses up to 150 times the average height (Pollehn, 1980). If this changes the pulse height distribution then the tartan may also appear. However, it may be prevented if the electronics discriminate against large pulses and temporarily inhibit address decoding.

The sum-of-all-masks is an artifact caused by the limitations of the image tube. The specifications of the intensifier, including the decay time of the phosphor and the type of microchannel plate, should be chosen to optimize the dynamic range of the camera.

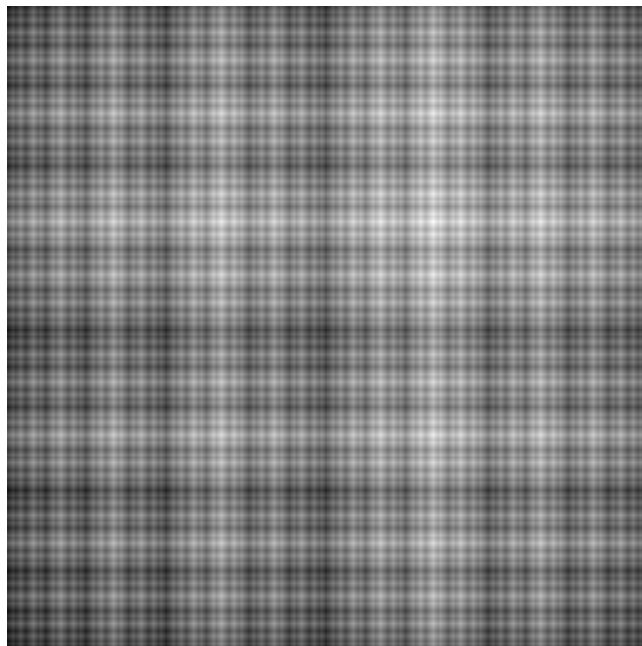


Figure 3.16: The sum of all masks. If all Gray coded masks were slightly transparent and placed one on top of the other, both x and y , then this is the image that would be produced. There are sixteen masks represented here for 8 bits of address information along each axis. The darker regions are where most address bits are off.

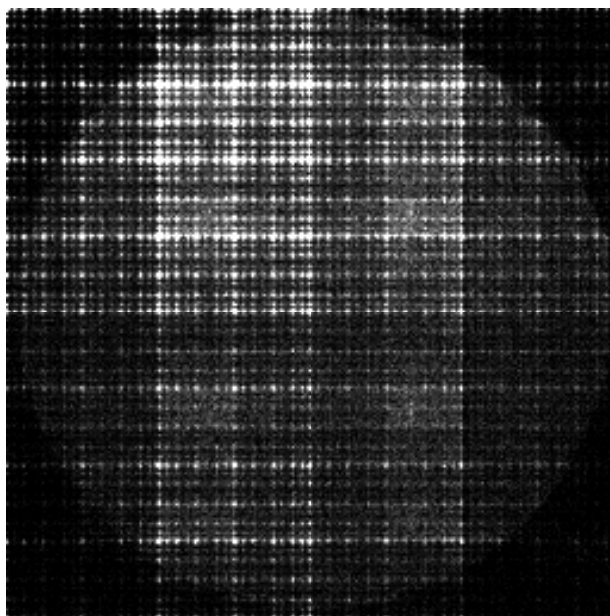


Figure 3.17: Light level problems. A much brighter flat-field was used for this image. The response of the image tube has changed so that aspects of the sum-of-all masks tartan are strongly superimposed here. Particularly noticeable are bright lines at addresses where most address bits are off.

Chapter 4

The Sydney University PAPA Camera

The design of the Sydney University camera will now be presented. Its new mask design will be described in detail, and the mechanical design of the new lens-mask assembly will be shown. The performance tests of the new detector are then discussed. Photographs of the completed camera are included in Appendix A.

4.1 Guidelines for the New Design

The new camera is based on the structural design of the Standley & Nisenson camera, which had been drafted in 1989 by Jeff Hazen at Harvard University. A complete set of the camera blue prints were obtained from Clive Standley and used as a basis for the rebuilding. The list of drawings is tabled in Appendix D. The main objectives were to correct for vignetting artifacts, reset the focal lengths of all array lenses, realign the optics of the camera, and also re-engineer and rebuild the analog electronics. Other changes were also performed to make the system easier to use. They are summarised below.

All major purchased components, with the exception of the mask plate, were retained for the new design. The image intensifier, cooler, collimating lens, array lenses, field lenses, and photomultiplier tubes were all kept.

The most important design goal was to obtain a new lens-array pattern to bring all lenses closer together and thus guard against vignetting. The choices were determined by purely geometrical constraints. Each of the Rolyn zoom lenses are 16.9 mm in diameter, and can be placed on centers 18 mm apart, while the large Isco lens has an effective diameter of only 93.75 mm. Array patterns were therefore discussed with those limitations in mind. Since there is only one pattern for 21 lenses, the one

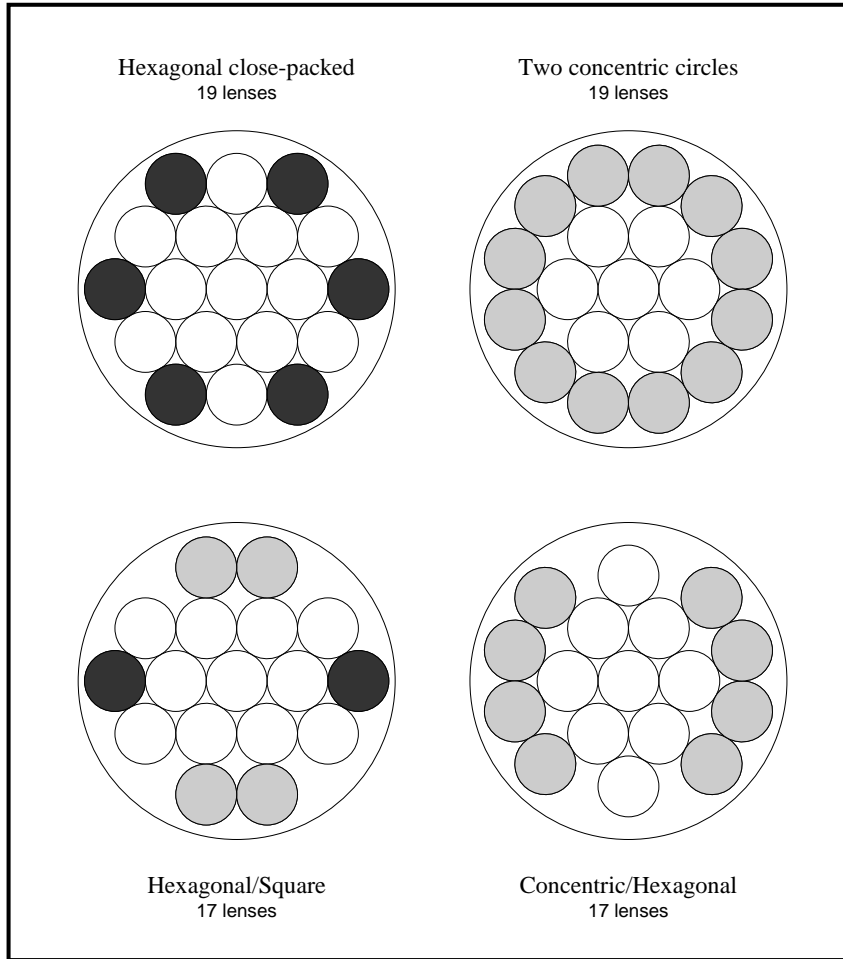


Figure 4.1: The 19 and 17 lens array patterns that were considered. The dark and light gray circles correspond to outliers for the 19-lens hexagonal and concentric patterns respectively.

used in the camera, it had to be abandoned. Alternate layouts for 19 and 17 lenses were examined.

Of concern was the magnitude of the vignetting and its orientation with respect to the masks. The orientation is important because if a vignetted region of the image falls entirely on an opaque part of a mask, then the vignetting does not bias the addresses produced by the camera. However, if the vignetting crosses several mask stripes then it will not be possible to flat field. It is therefore possible to use the placement of the lowest order masks—masks which are largely opaque—as a shield against address errors produced by vignetting. However, this would only be effective with arrays that have the appropriate symmetry.

The four array patterns that were considered are shown in Fig. 4.1. The smaller circles represent the 18 mm diameter spacing of the array lenses, and the larger circle represents the 93.75 mm effective aperture of the main lens. There are two 19-lens

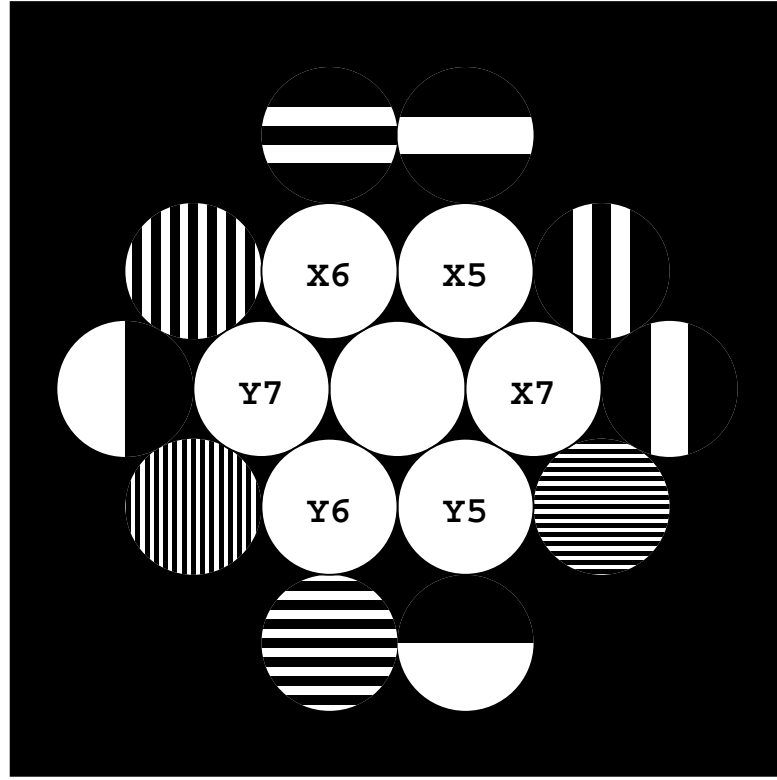


Figure 4.2: The Gray coded mask plate. All the masks are shown except the higher order masks surrounding the strobe channel. The strobe channel itself is transparent, as indicated.

patterns, and two 17-lens patterns. The furthest lenses in the hexagonal pattern are its 6 corners, which have been shaded dark gray. The furthest lenses in the pattern of two concentric circles—12 lenses all equidistant from the center—have been shaded in light gray. The two other patterns, with 17 lenses rather than 19, are combinations of the first two patterns. In each case the furthest lenses have been shaded to reflect the symmetry of the 19-lens patterns from which they were borrowed.

The 17 lens array that was chosen, the Hexagonal-Square pattern, has several advantages over the others. The worst-case lenses are bunched near the x and y axes; the two hexagonal (dark gray) outliers lie on the x axis, whereas the four concentric (light gray) outliers are bunched close to the y axis. In every other pattern the worst-case lenses are distributed more evenly and therefore less advantageously. Moreover, only 6 lenses lie in locations likely to suffer from vignetting, and only two of those are as bad as the worst-case hexagonal lenses. The hexagonal pattern also has only 6 outliers, but these are all equidistant from the center, and suffer from vignetting equally—without the advantages of a useful symmetry.

It was thought that the few outliers and the symmetry offered by the Hexagonal-Square pattern would offer the best chance of alleviating the vignetting problem, with

the restriction, of course, that a 17 lens array must be adopted.

4.2 The Mask Plate

The detailed design of the new mask plate is given in Appendix B. It is also shown in full scale in Fig. 4.2. The lower order masks have been positioned to take advantage of the symmetry, as discussed previously. The vignetting would cause the images to have a tapered illumination on the side of the mask nearest the center. The outer masks are therefore oriented so that their opaque regions lie towards the strobe channel; x masks are placed left and right, and y masks are placed top and bottom. The X0 and X1 masks have been placed opposite each other on the central row, lying along the x axis. The Y0, Y1, Y2, and Y3 masks have been bunched along the y axis. The Y3 mask would be the worst to suffer from vignetting, since it does not have large regions which are solidly opaque. However, there is no evidence of vignetting in the images.

4.3 The Mechanical Design

The geometry of the new mask plate required that the lens-mask assembly and the photomultiplier housing be redesigned. The new drawings for these components are shown in Figs. 4.3, 4.5, and 4.4. These were based partly on the previous drawings by Jeff Hazen, noted in Appendix D.

4.3.1 The lens-mask assembly

The array lenses were mounted on barrels and slid into a mounting plate of Fig. 4.3. This arrangement was used so that with each lens the adjustment of focus and rotation was independent. The rotation is required to align the image on the masks, as was discussed in §3.2.3. As depicted in the figure, the plate allows lateral access for focus and rotation to all lenses but the central three: the strobe and the finest x and y masks. Those may be fixed in place without compromising the alignment; the lenses of the two finest masks serve as the reference for the alignment of the x and y axes, and may be fixed in place providing they are focused correctly; the rotation and focusing of the central strobe is not crucial, as it serves only to collect light.

In the previous design the alignment had been done using tilted plates of glass, 1 mm thick, located between each array lens and its mask. The lenses were held fixed and the tilt-plates were rotated. The tilt-plates were placed at an angle of 30° to the converging beam, and allowed each image to be adjusted $\pm 200 \mu\text{m}$ on the surface of the mask. However, it was found that they alone did not ensure an alignment would be possible: some of the lenses needed to be rotated as well. However, the lenses

had been threaded into their mounts, so that rotating them caused their images to defocus. These problems were corrected with the new design.

Now that the rotation and focus adjustments were made independent, it was found that the tilt plates were no longer necessary, as the required range of adjustment was available from the lenses alone. Also, because each lens was now held in place by a screw adjustment, the mechanical stability and ease of alignment were greatly improved.

The new mask plate, supported on three ball bearings, is held in place by the holder depicted in Fig. 4.4. Three hex screws, cushioned by cork pads, were tightened onto the glass directly above the ball bearings. Invar rods were used to separate the plate holder from the lens holder of Fig. 4.3. These were attached to each piece by screws which ran through the holders at the 0.281 inch diameter holes.

The support of the lens-mask assembly was also altered. The assembly was brought closer to the collimating lens by shortening the main support struts of the camera's base. This was another measure to guard against vignetting by that lens. Also, the mounting plate that holds the lens-mask assembly was redesigned to allow better access to the array lenses and permit the alignment to occur in situ. With the new mounting plate the invar rods of the assembly were in line with the support struts. They had blocked direct access to the lenses in the previous design. The mounting plate is shown in Fig. 4.6.

4.3.2 Mounting of the field lenses

The vignetting by the field lenses, noted in §3.4.4, was corrected as well. The lenses were mounted in the front of the photomultiplier tube housing shown in Fig. 4.5. The holes through to the PMTs were made larger, and the lens mounts were brought closer to the mask plate.

4.3.3 The realignment of the optics of the camera

With the help of Derek McKay the new lens-mask assembly was optically aligned. The Rolyn zoom lenses were removed from the previous camera, taken apart, and cleaned. They were then mounted in the lens-mask assembly and were adjusted in situ as described in Appendix C.

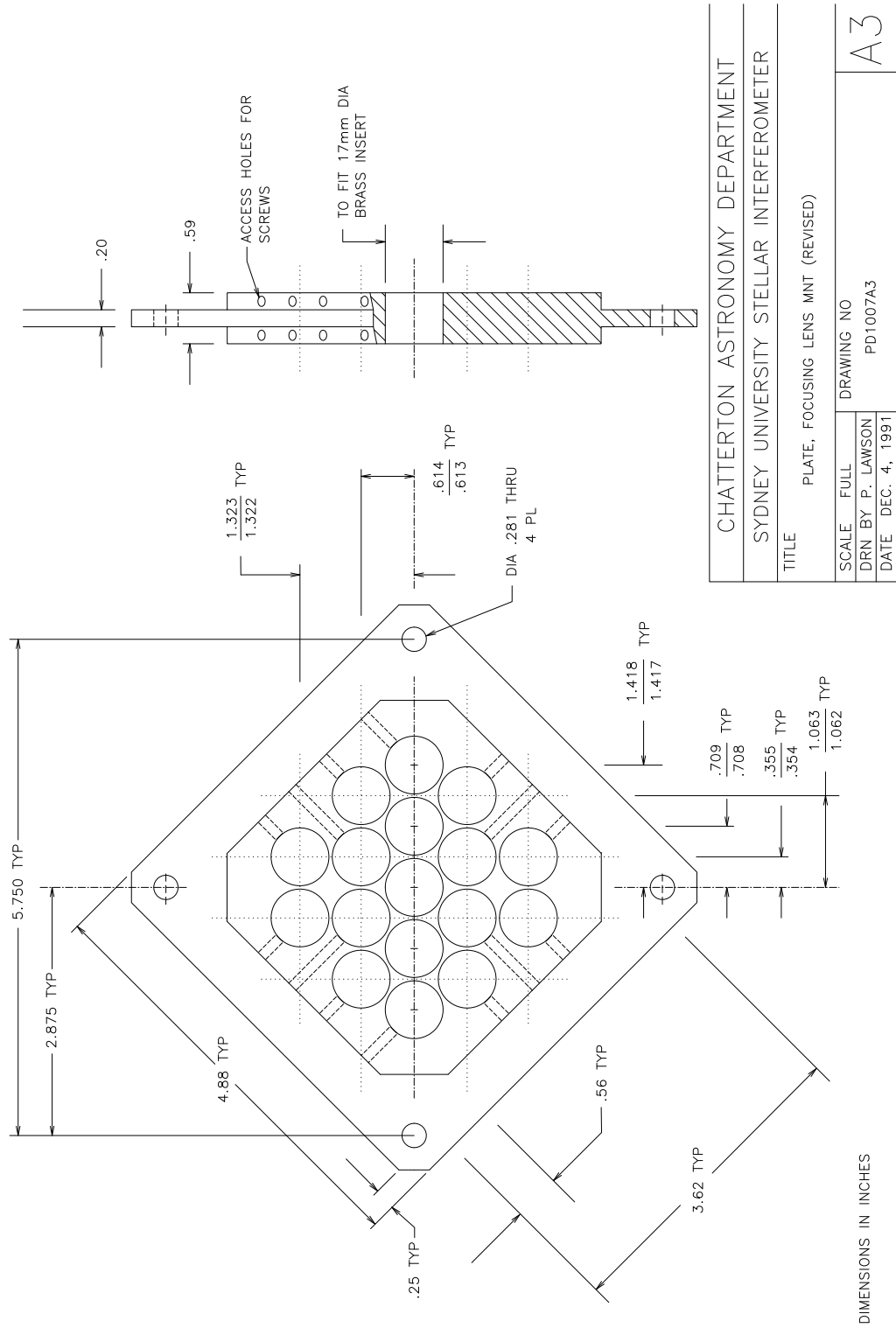


Figure 4.3: The mount for the array of lenses. This would also have held the tilt-plates had they been necessary.

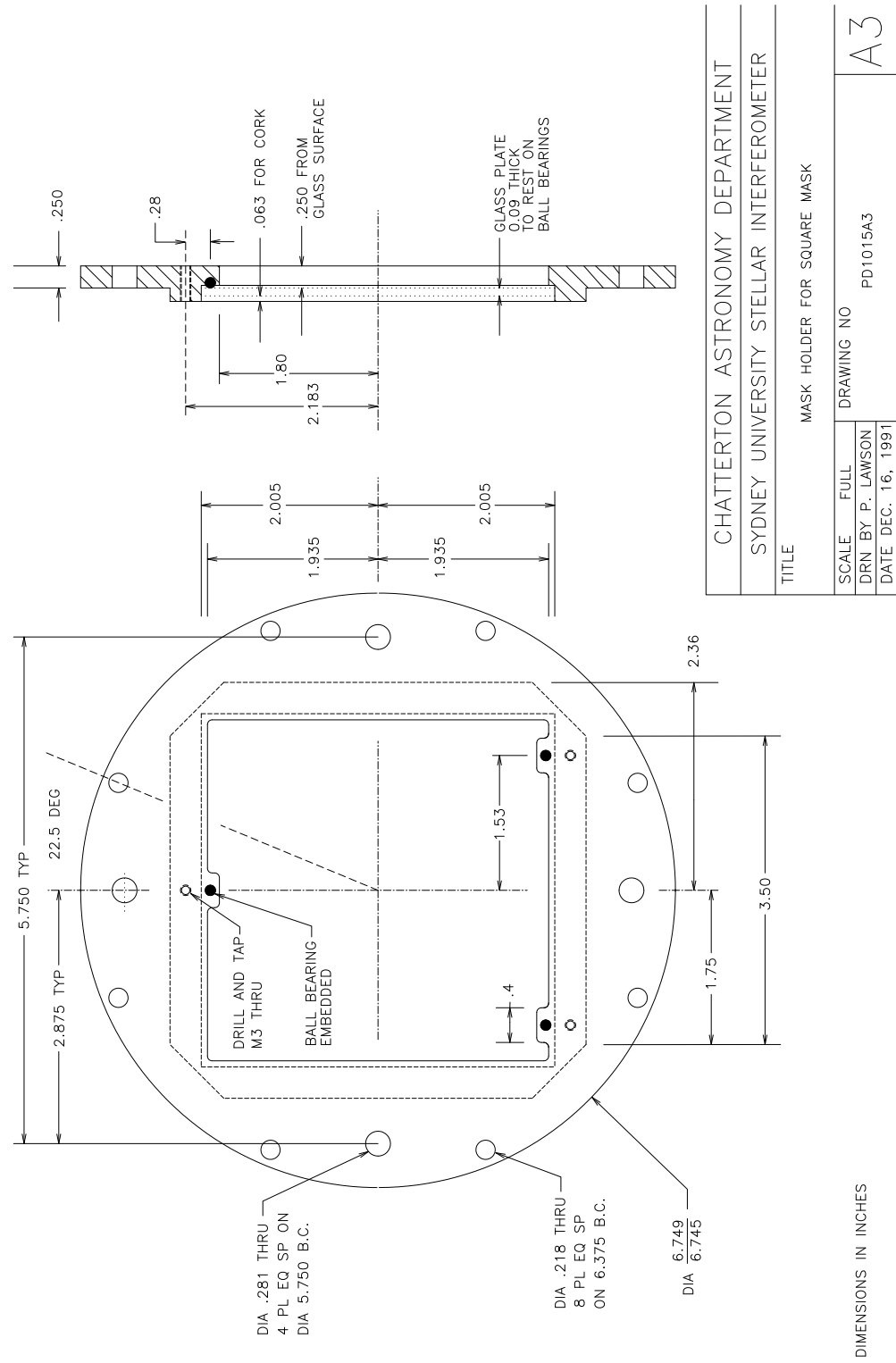


Figure 4.4: The support for the mask plate.

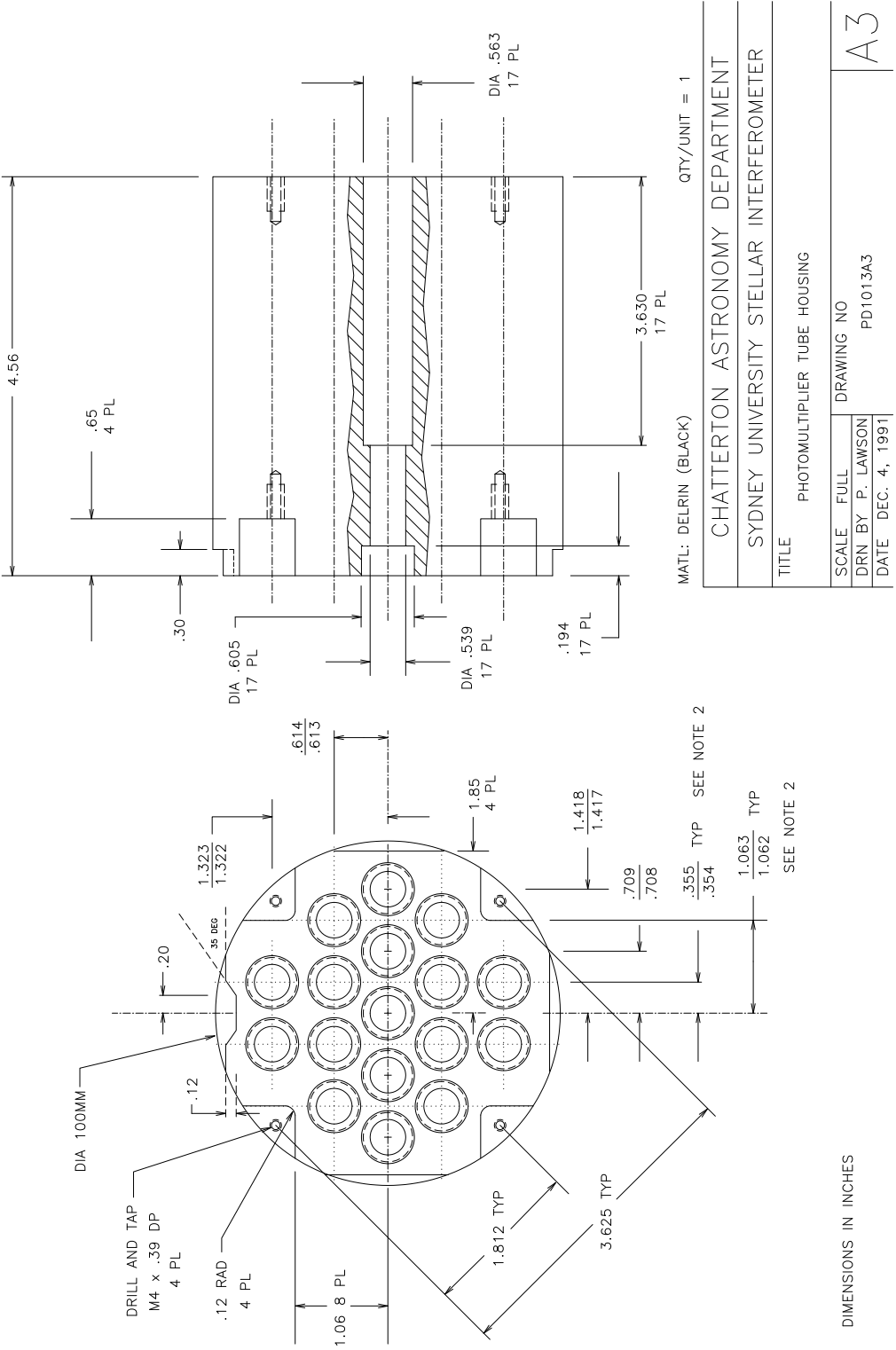


Figure 4.5: The photomultiplier tube housing.

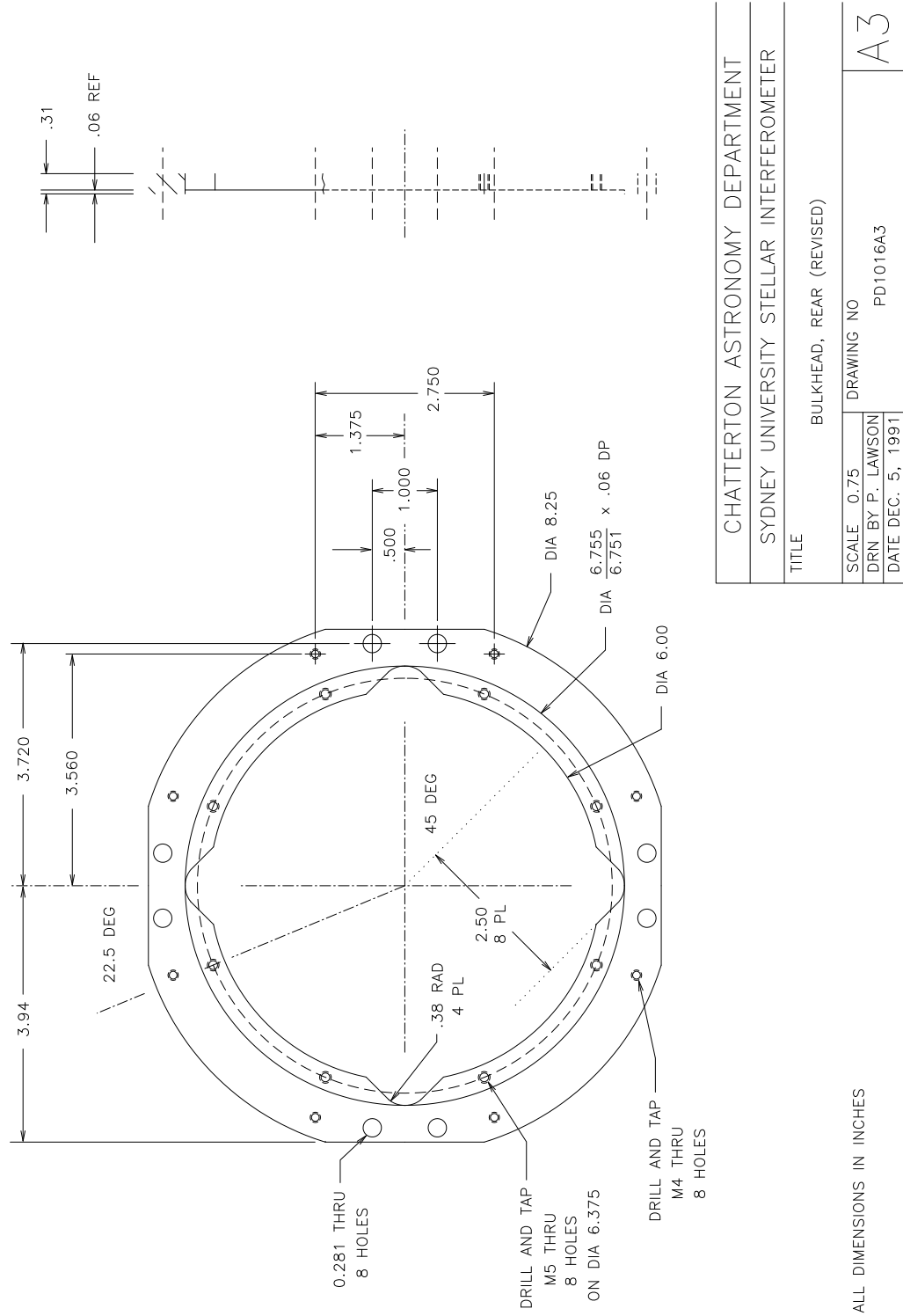


Figure 4.6: The mounting plate for the lens-mask assembly. The support plate for the mask fits into the center of this piece.

4.4 The New Analog Electronics

The analog electronics were rebuilt to guard against electrical noise. New printed circuit boards were designed and layed out by Hank Bennis and Derek McKay in consultation with the author and following the suggestions by Cañas (1989). Most of the changes were to improve the engineering. There were three printed circuit boards in the new design: two identical yet separate x and y boards, and a strobe board. Included were voltage regulators at each board and a ground plane on the board's rear surface. Miniature BNC connectors were used to connect the coaxial cables from the photomultiplier tubes. The circuitry of the strobe board was substantially changed.

A cooler was also designed and built by Fred Peterson to enclose the analog electronics. The intention was not simply to cool the electronics, but to remove the heat they generated from the enclosure of the interferometer.

4.5 Software and Computer Interface

Software was written in the C programming language to interface the camera to an IBM AT clone. The interface board was a 32-bit parallel card, the National Instruments AT-DIO-32F, capable of Direct Memory Access (DMA) and acquisition speeds up to 450 K bytes per seconds (National, 1990). The acquisition was driven by DMA software routines described by Nolan (1990).

The software allowed the collection of images, reading and writing data files, and could graph cross-sectional views of the data with overlaid mask boundaries—a useful diagnostic. It performed the calculations required to adjust discriminator settings and could also translate images into PostScript format. All PAPA camera images displayed in this thesis were made using this software.

4.6 Performance Tests

The performance of the camera will now be evaluated by examining its flat-field response, resolution, sensitivity, linearity, and its noise properties.

4.6.1 Flat field

The flat field response was previously presented in the discussion of image artifacts. It is shown in Fig. 3.9. The cross sections of Figs. 3.10 and 3.11 indicate in greater detail the success of the alignment. The overall appearance of the flat-field is good,

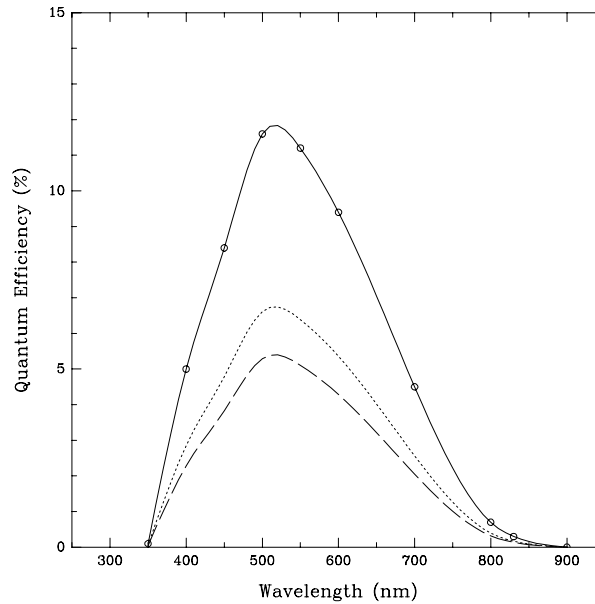


Figure 4.7: The quantum efficiency of the photocathode of the Varo Gen II intensifier, Type 3036, Serial No. 907026. The open circles indicate the measurements made by the manufacturer. The quantum efficiency is derived from measurements of the photoresponse (mA/W), the current generated by the photocathode. This current is measured with the front and back surfaces of the MCP shorted to ground. The detected quantum efficiency of the image tube is therefore almost half this value, since the open-area ratio of the MCP is only 57%. The dotted line is 57% the response of the solid line.

with rms pixel-to-pixel variations at around 10% of the mean. This suggests that, on average, the mask edges were aligned to within a tenth of a pixel's width, $\beta = 0.1$. These variations in the effective size of pixels account for most of the structure in the flat-field.

Several larger errors are also present where the mask alignment has slipped by more than one pixel. These are due to chromatic aberration in the zoom lenses, resulting in a field dependent defocus. Defocusing is evident in the edge of the X1 mask, which is poorly defined: the width of the feature there will broaden according to the discriminator threshold, and may be improved if the threshold is changed by slightly less than 1%.

4.6.2 Detective quantum efficiency

The Detective Quantum Efficiency (DQE) of the camera is limited by the response of the image tube's photocathode. This is plotted in Fig. 4.7. As was discussed in §3.3.1 the DQE of the intensifier will be approximately 57% of this, limited by the open-area ratio of the microchannel plate. If the efficiency of the rest of the camera were 80%

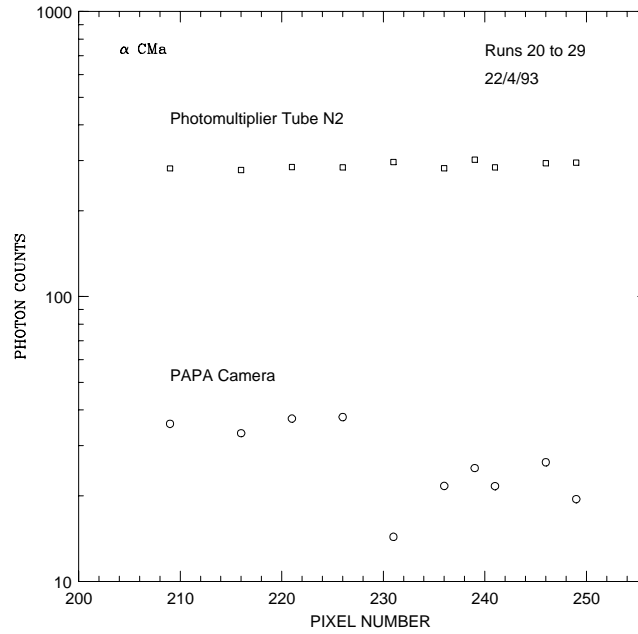


Figure 4.8: A comparison between counts from the N₂ photomultiplier and selected pixels on the PAPA camera. The PMT measures counts from a 0.4 nm bandpass centered on 441.6 nm. The bandpass of each pixel is the same, but may be centered on different wavelengths. The PAPA camera response has been scaled to include the 20:7 ratio in aperture diameters.

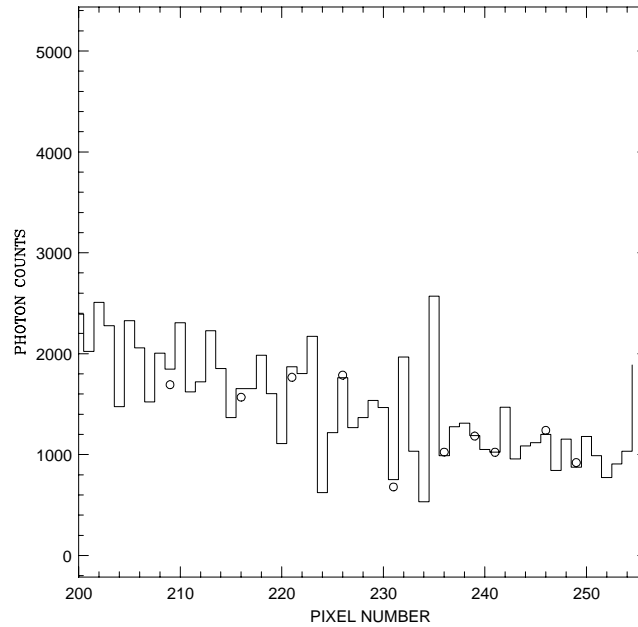


Figure 4.9: An enlargement of the spectrum of α CMa as measured with the PAPA camera software and 32-bit parallel interface. Overlaid on the histogram are the PAPA camera data points shown in the previous figure.

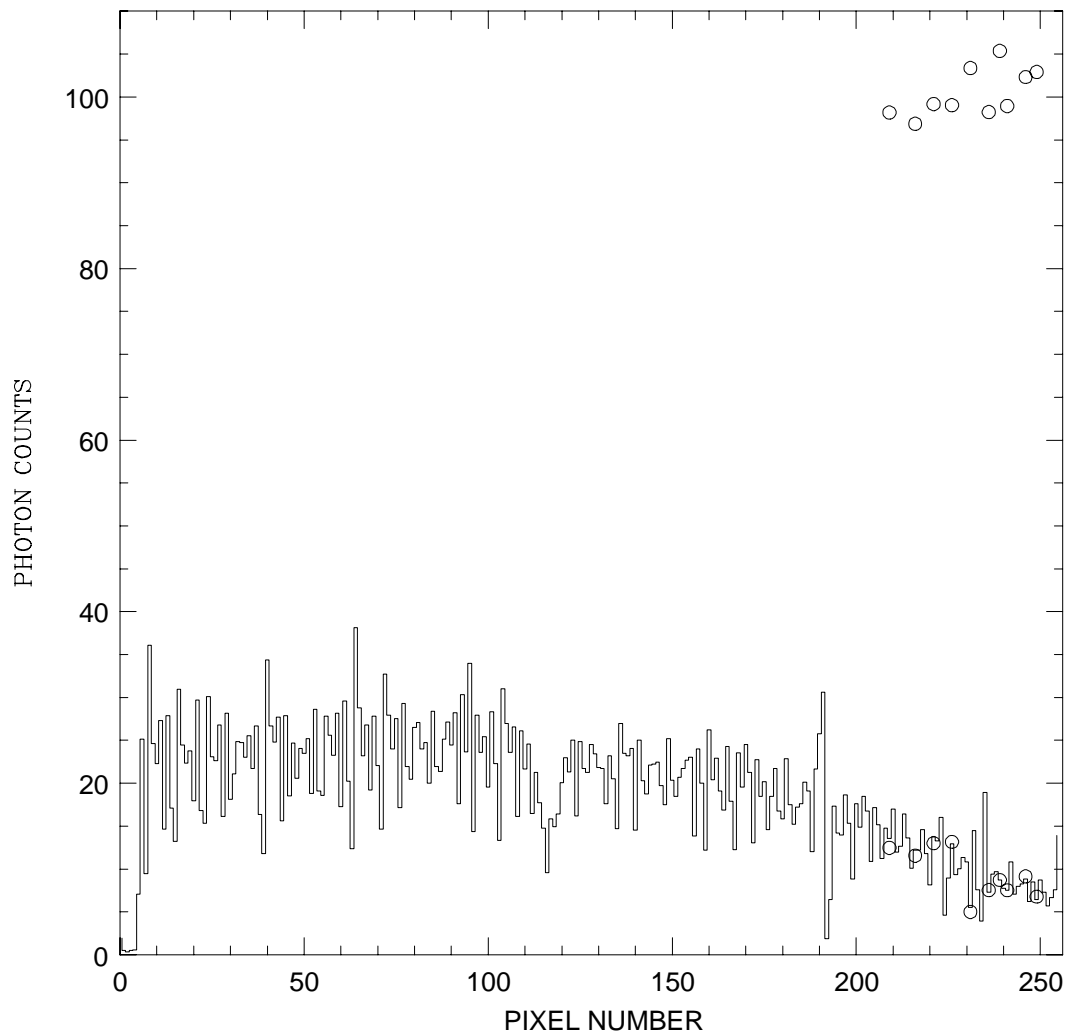


Figure 4.10: The full spectrum of α CMa with the data of April 22. The spectrum has been scaled to fit the response of the PAPA camera measured relative to the photomultiplier tube data. The open circles are from the same data that was shown in the previous figures. It can be seen that in the blue, towards higher pixel numbers, the camera has a quantum efficiency of approximately 10% of the photomultiplier tubes. A broad absorption line is visible in the middle of the spectrum.

then the DQE of the system would now be estimated by the dashed line in the figure. At a wavelength of 440 nm it would be approximately 3.5 %. Measurements on site at the stellar interferometer indicate that the DQE is probably closer to 2.0%.

The DQE was estimated through comparisons with the sensitivity of the Photomultiplier Tubes (PMTs) at the Sydney University Stellar Interferometer. These are shown in Fig. 1.1. After the light has passed through the beam combiners it is normally sent to the N_1 and N_2 PMTs. However, light going to N_1 was redirected to the PAPA camera using a periscope. With this configuration the light was equally divided between the camera and N_2 , so direct comparisons were possible. The camera was used as a detector in a low dispersion spectrograph, using 430–550 nm, across 256 pixels. Red light was dispersed to low pixel numbers, and blue light to higher ones. Each pixel represented a bandwidth of about 0.4 nm.

Observations of Sirius, α CMa, were conducted on April 22, 1993. The N_2 PMT was set for a bandwidth of 0.4 nm to correspond with the bandwidth of a single pixel on the camera. A 20 mm diameter aperture was used for the combined beam going to the camera, and a 7 mm aperture was used for the beam going to N_2 . This produced about 280 counts per second in each channel. Individual pixels on the camera were selected for this comparison. These corresponded to different neighboring wavelengths. The results for pixels between numbers 200 and 250 are shown in Fig. 4.8, and a spectrum of α CMa is shown in Fig. 4.9. In each of these plots the open circles represents the same data. The spectrum was taken separately using a longer exposure time, it is drawn as a histogram, and was rescaled to fit the data. When the full spectrum is rescaled to the PMT response in Fig. 4.10, it can be seen that the sensitivity varies from 10% to 20% of that of the PMT. If the PMT has a DQE of 20% then the camera would have a response of near 2%.

On April 4 the counts across the whole camera were compared in the same way with N_2 . The bandwidth of the photomultiplier tube was set at 0.75 nm and 10 mm apertures were used for both combined beams (Run Number 00015). The mean counts were 24726 per second recorded at the camera as compared with 1038 at the photomultiplier. The estimated quantum efficiency of the detector is then 18% of the PMT. This is in accord with the previous measurements.

4.6.3 Resolution

The limiting resolution of the camera can be seen in the images of Figs. 4.11 and 4.12. The charts were illuminated by a red LED and imaged onto the camera with a 230 mm focal length achromat, for a count rate of near 500k photons per second. Slightly higher light levels were used to focus the system so that the finest lines were visually resolved on the output of the image intensifier. Figure 4.11 shows an image of a resolution chart, and Fig. 4.12 shows the image of a fine grid of fine lines. 97 line

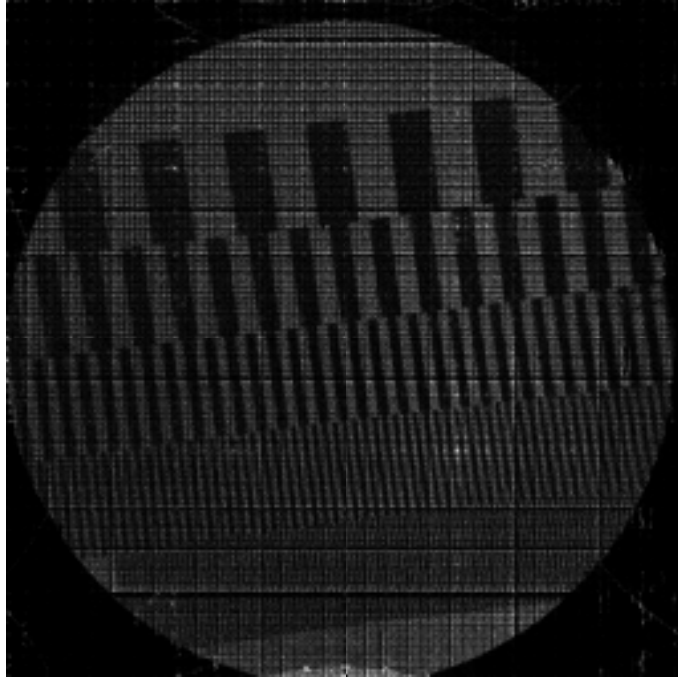


Figure 4.11: An image of a resolution chart taken at a data rate of 500k photons per second. The finest grid was resolved by the image intensifier, but is not fully resolved in this image. It represents line pairs where the white and dark lines each span one pixel. The target image was intentionally tilted so that the resolved lines would not be confused with fine scale artifacts.

pairs are resolved in this figure, indicating that at least 194 pixels are present. In each figure there is also a moire pattern across the grids at the highest resolution, indicating a beat between the grid pattern and the pixel dimensions. This is consistent with a resolution approaching 256 pixels along each axis. An example image is shown in Fig. 4.14.

4.6.4 Linearity and dark count

The linearity of the camera is shown in Fig. 4.13. This measurement was done using a set of neutral density filters from the Schott glass catalog. A red light-emitting diode was used as a source to illuminate a white sheet of paper, which was then imaged onto the camera, with a Hewlett Packard pulse counter used to monitor the data rate. It can be seen from this diagram that the camera has a linear response over a large range. At count rates higher than 1 million photons per second the sum-of-all-masks, described in §3.5.4, begins to appear in the image. A dark count of 200 photons per second is indicated by this graph.

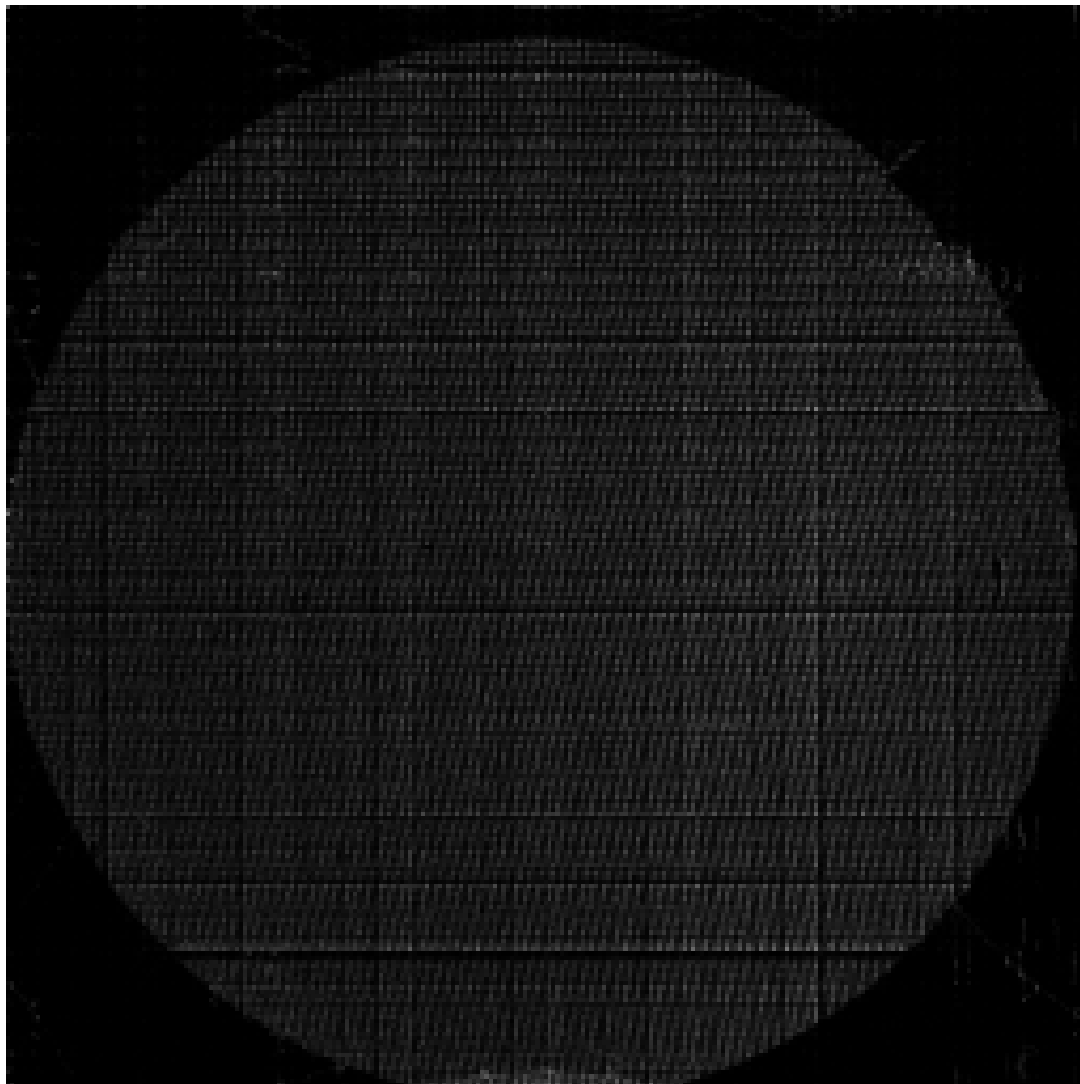


Figure 4.12: This is an image of a field of finely spaced line pairs which have been tilted about 20 degrees from the vertical. There are 97 line pairs resolvable across the diameter, indicating a resolution of at least 194 equivalent pixels.

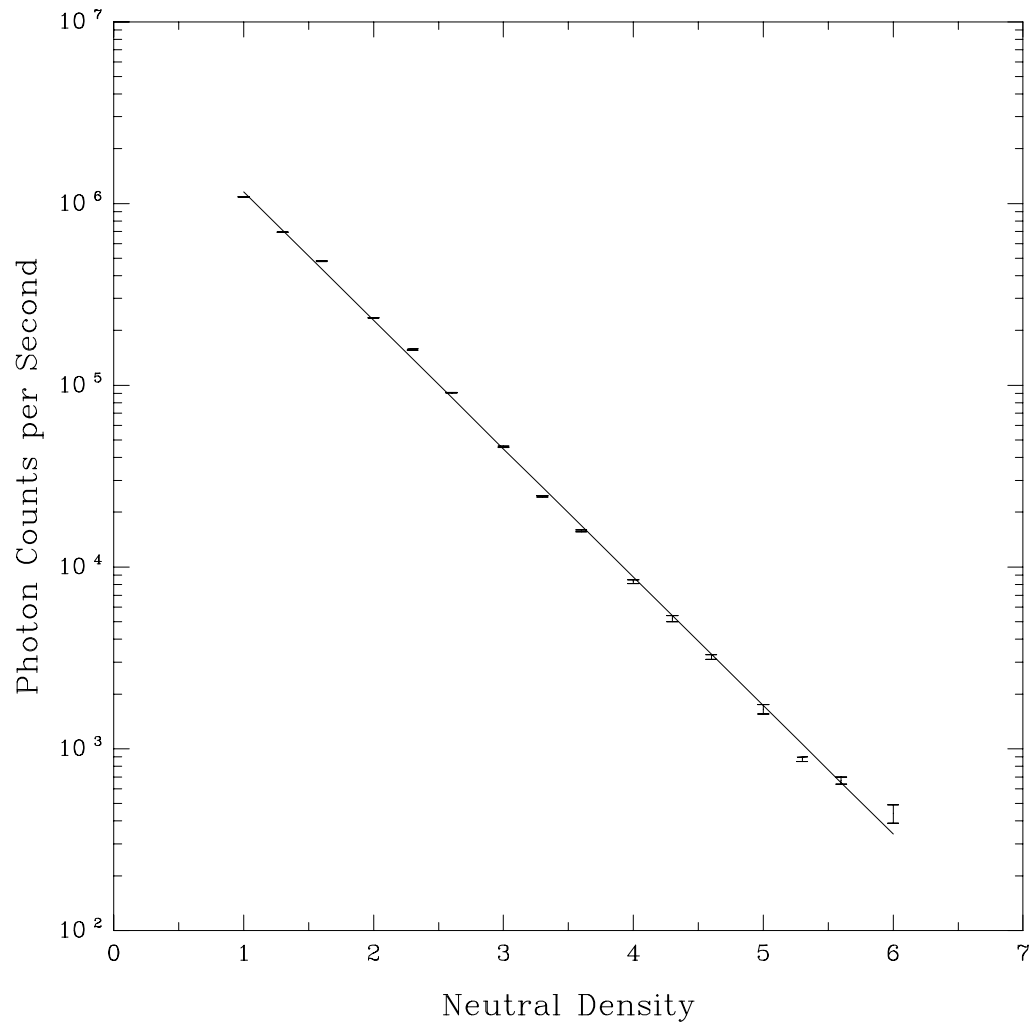


Figure 4.13: The linearity of the camera. The number of counts per second is shown as a function of neutral density. A red LED was mounted in an enclosure and used to illuminate a white sheet of paper, which was then imaged onto the camera. The background counts are at a level of 200 counts per second. It can be seen that the camera is linear over a very wide range of illuminations. High light-level artifacts begin to be seen at count rates greater than 1 million counts per second.



Figure 4.14: The logo of the Sydney University Stellar Interferometer imaged with the PAPA camera.

4.7 Summary

In Chapt. 3 the theory and operation of PAPA cameras were presented, and the sources of image artifacts were described in detail. The Sydney University camera was developed around a new mask plate, and was specifically designed to prevent the occurrence of vignetting artifacts. It was shown in this chapter that the new camera achieved that objective. Other changes that were included were a new housing for the photomultiplier tubes, and a new mount for the array lenses. The most challenging aspect of the optical design arose from the restrictions placed on the array lenses. Each lens must have its effective focal length matched to better than 0.15% over the spectral response of the image tube phosphor. They must all be mounted as close together as possible, be independently adjusted for focus and rotation, and locked in place after alignment. The array pattern of the mask-plate allowed the zoom lenses to be packed within the diameter of the collimating lens while allowing lateral access to each lens for independent adjustments of focus and rotation. This feature was a significant improvement over the previous design because it allowed the alignment to be performed without the use of tilt-plates.

The flat field that was depicted in Fig. 3.9 is a typically good response for the optics of the Sydney University camera. No vignetting artifacts are visible. The residual alignment and magnification errors are due to chromatic aberration and image

distortion produced by the small lenses used in the array.

The camera was shown to be able to resolve at least 97 line pairs across its field, indicating resolution on a scale of 194 pixels. The camera has a linear response for count rates up to 1 million photons per second, and a dark count of only 200 photons per second over a full field of 256 x 256 pixels. Although direct measurements of detective quantum efficiency were not made, it was estimated to be near 2% at 440 nm, based on a comparative observation of stellar sources between the camera and the photomultipliers of the stellar interferometer. This is comparable with the sensitivity of previous PAPA cameras; a DQE of 3% was quoted by Papaliolios et al. (1985), and a DQE of 0.5% was estimated by Mozurkewich (1993) for the Adaptive Optics Associates camera at the Mark III interferometer. It is also comparable with other photon counting cameras. Morgan (1988) states that an S-20 MAMA camera has a DQE of 4%, whereas a previous model with a bialkali photocathode had a DQE of only 0.5%.

Suggestions for future improvements to the Sydney camera are given in the concluding chapter, Chapt. 12. In Part III we now return to the main topic of the thesis and provide background for the experiments which are to follow.

Part III

Theory of Group Delay Tracking

Chapter 5

Delay Tracking with the FFT

Part III of this thesis is concerned with the theory and limitations of using the Fast Fourier Transform for group delay tracking. It is now appropriate to review this approach and to introduce the work that will be covered in the following chapters. The theory precedes the experimental work which is described in Part IV.

5.1 The Channeled Spectrum

The performance of the FFT will be predicted based on the simple model of the channeled spectrum which was introduced in Chapt. 2. Although the model provides an optimistic estimate of the performance it is nonetheless a useful benchmark. It is as follows:

$$I(\lambda) = I_s \left[1 + |\gamma| \cos \left(\frac{2\pi x}{\lambda} - \alpha \right) \right] + I_b, \quad (5.1)$$

where λ is the wavelength of light, I_s is the intensity of the unmodulated stellar spectrum, I_b is the background intensity, x is the optical pathlength difference, and γ is the complex fringe visibility. This is identical to Eq. 2.1.

Bright fringes appear in the spectrum wherever the path difference is an integral number of wavelengths. The number of fringes across a fixed bandwidth increases with increasing delay, and so the path difference can be monitored by observing the number of fringes. If n fringes are counted in the spectrum between wavelengths λ_{min} and λ_{max} , then it is straightforward to determine the pathlength x . From Eq. 5.1 it can be seen that

$$n = \frac{x}{\lambda_{min}} - \frac{x}{\lambda_{max}},$$

and therefore

$$x = n \left[\frac{1}{\lambda_{min}} - \frac{1}{\lambda_{max}} \right]^{-1}.$$

The number of fringes is a linear function of the path difference.

5.2 Periodogram Estimate of the Delay

The delay may be found by determining the frequency of the fringes measured in units of ‘fringes per interval of $1/\lambda$ ’. Figure 5.1 shows a channeled spectrum and its corresponding Fourier transform. The spectrum is dispersed so that the fringes lie at regular intervals in the detector coordinate. If the detector had unlimited resolution, and was infinite in extent, then its Fourier transform would contain a dc component and delta functions at \pm the fringe frequency. In this example the spatial frequency is the number of fringes per 256 pixels, that is to say 23.

For real valued data, such as photon counts, the positive and negative frequency components are complex conjugates. No information is neglected if the negative frequency components are ignored. Two further examples are given in Fig. 5.2 where now only positive frequencies components are displayed. In this case a Discrete Fourier Transform was used to identify the fringe. The detector now has a limited number of pixels, and the fringe response is no longer a delta function. Under practical conditions the situation is complicated even further. In general, most parameters in Eq. 5.1 will vary with wavelength. The optical pathlength x will be wavelength dependent and modified by the longitudinal dispersion in the two arms of the interferometer. The spectrum and the background will have structure, and the visibility of the observed source will change as a function of wavelength. As well, the spectrum will be represented by discrete photon events and at low light levels will be sparsely sampled. The pathlength difference may also change during the observations and cause the visibility of the detected fringes to be reduced.

Moreover, the process of detection alters the appearance of the fringes. The stellar spectrum is truncated and smoothed by the pixels of the detector, and their finite size sets the coherence length for the observation: small pixels represent narrow bandwidths and therefore large coherence lengths. The mapping of the fringes onto the detector, determined by the angular dispersion of the spectrometer, may also reduce the response to large delays. The dispersion is usually defined as

$$\mathcal{D}_\lambda = \frac{d\epsilon}{d\lambda},$$

where ϵ is the angle of deviation of the light (Born and Wolf, 1980). However, in the following discussions the term dispersion will be used to denote the change in ϵ as a function of spectroscopic wavenumber, $\kappa = 1/\lambda$.

$$\mathcal{D}_\kappa = \frac{d\epsilon}{d\kappa},$$

By this definition, if ϵ is a linear function of wavenumber, $\epsilon(\kappa) = c_0 + c_1\kappa$, then the dispersion is constant and the channeled fringes will appear at regular intervals in the detector coordinate. If the dispersion is non-constant then the spacing between fringes will change and the fringes may be stretched or compressed.

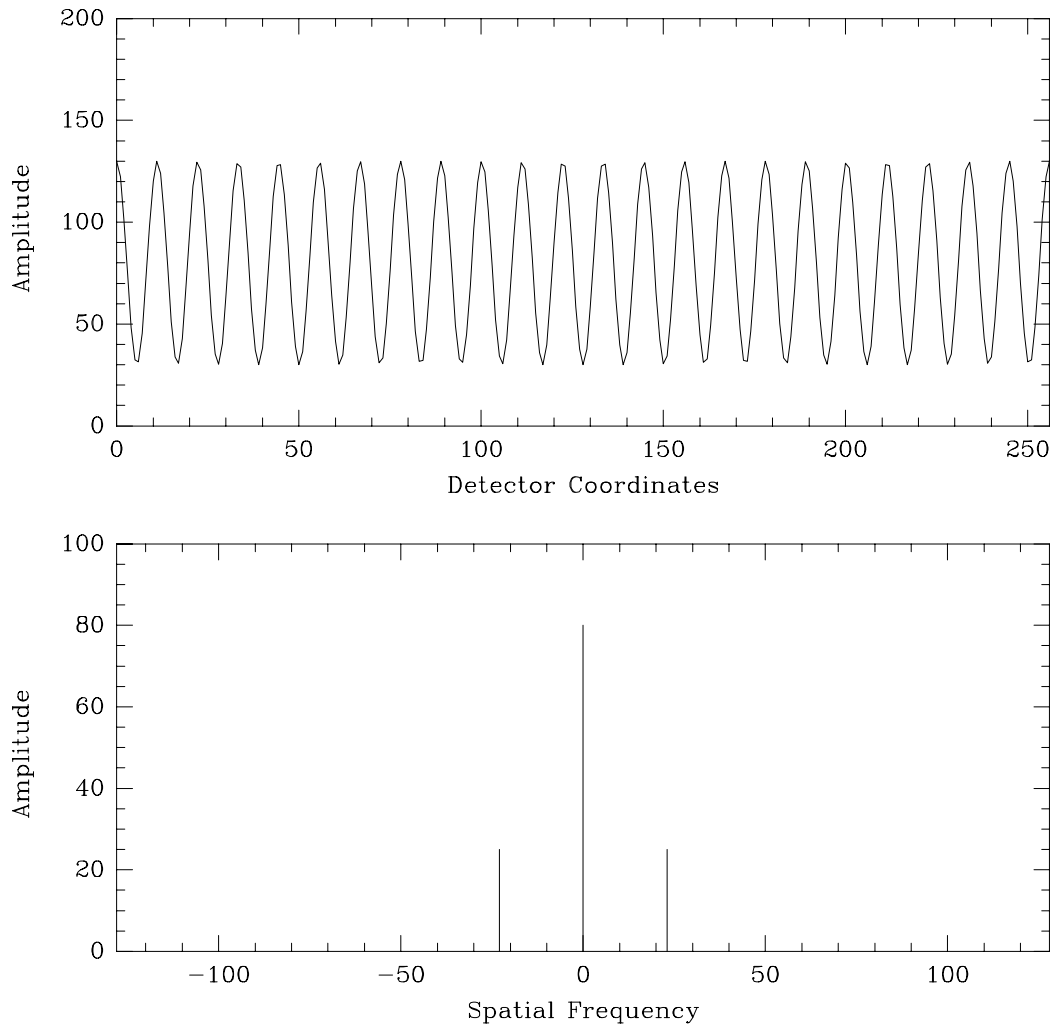


Figure 5.1: A model for the fringes. In this example there are 23.0 fringes with constant spacing across the detector. The fringes have an amplitude of $I_0 = 50.0$, a background level $I_b = 30.0$, and visibility $|\gamma| = 1.0$. The Discrete Fourier Transform shows a zero frequency level of $I_0 + I_b = 80$, and two components at spatial frequencies of ± 23.0 , each with an amplitude of $I_0|\gamma|/2 = 25.0$.

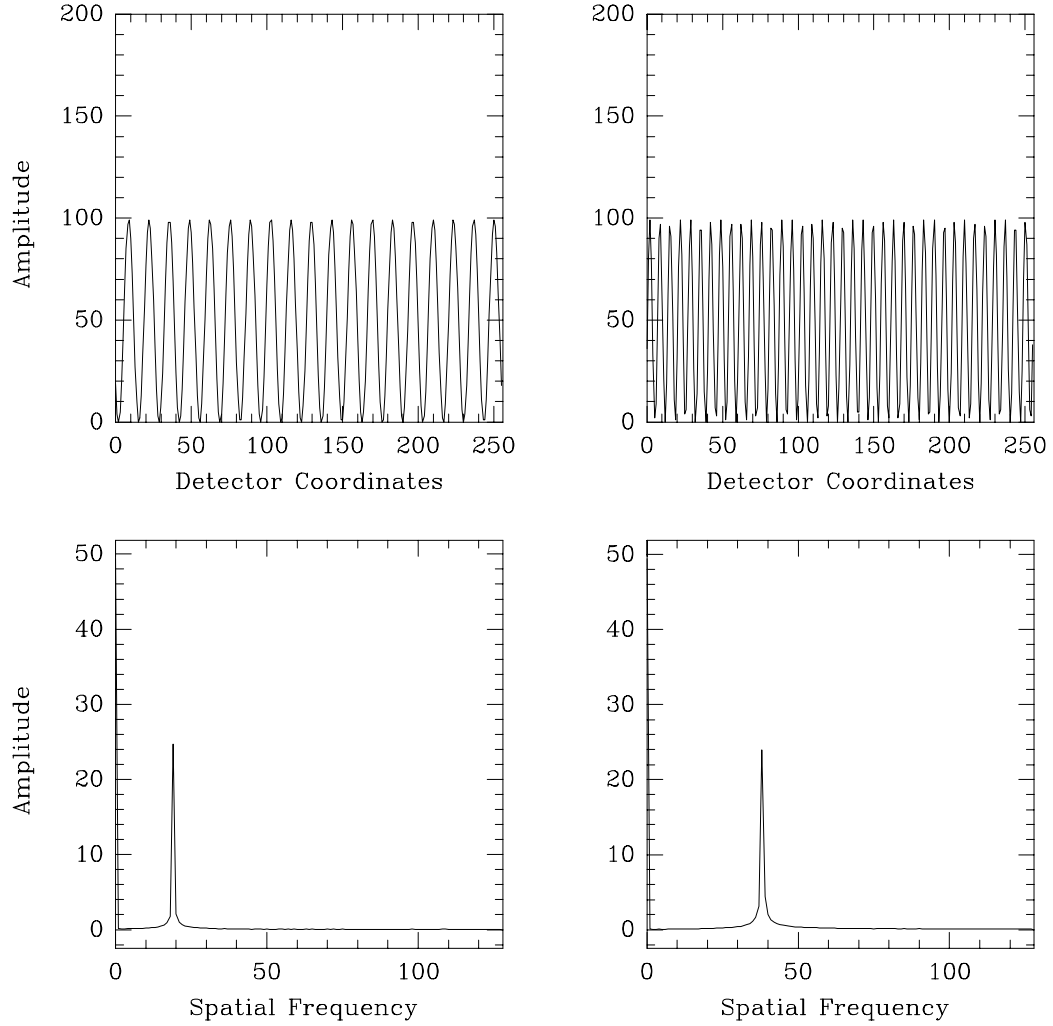


Figure 5.2: Spatial frequency response with dispersion constant in wavenumber and a limited number of pixels. Two examples are shown of fringes, top, and their Discrete Fourier Transform, bottom. The peak in the spatial frequency spectrum is located at the fringe frequency. It identifies the number of fringes counted across the detector, in this case across 256 pixels.

All these phenomena have corresponding effects on the Fourier transform and will be discussed in the following chapters.

5.3 Outline of Part III

Chapter 6 is concerned with how the process of detection limits the available information in a channeled spectrum. The fringes are sampled by an array detector, whose transfer function alters the transformed data. In §6.1 it is shown that this is caused by the sampling and the use of a discrete transform. This section should be read with reference to Appendix F, where the relationship between the Discrete Fourier Transform and the integral Fourier transform is discussed. Sampling the fringes introduces a coherence length, which limits the largest observable delay and sets the resolution by which delays can be estimated. Some more obscure artifacts are examined, and the use of zero-padding for interpolation is discussed. In §6.2 the effect of a non-constant angular dispersion in the spectrometer is considered. This causes the fringes to be compressed or stretched in the detector coordinate, spreading the spatial frequencies needed to describe the fringes. The Fourier components no longer lie at a well defined locations and the peak height is lowered at the nominal fringe frequency. It is shown that by means of a lookup table the photon data can be rebinned to avoid losses. The response then returns to nearly a delta function, at the expense of some noise at high spatial frequencies.

Chapt. 7 describes the low light limitations of detecting channeled spectra. The performance is limited by the discrete nature of the photon events which are assumed to obey Poisson statistics. Two approaches are used to describe the signal-to-noise ratio of the spectrum.

1. The first in §7.2 considers the squared modulus of the transform to be the parameter of interest—the power spectrum. It provides an unbiased estimate of fringe parameters, and its signal-to-noise ratio has been extensively reviewed in discussions of speckle interferometry. The method of derivation is reviewed and the results are presented.
2. The second approach in §7.3 examines the modulus of the transform as the parameter of interest. This is of importance because the modulus is integrated in the experiments discussed in Part IV. The analysis would be the same as that used for fringe location in radio VLBI (Very Long Baseline Interferometry), but the assumption needs to be made that the noise in the transform is Gaussian distributed—an unreasonable assumption under conditions of low signal-to-noise. Approximations are therefore used in the second approach.

In §7.4 it is shown that there exist better performance indicators than the signal-to-

noise ratio. The probability of tracking error is therefore derived and is compared with predictions from theory. This is a problem related to signal-to-noise issues, but one that has not been addressed in the literature. A figure of merit is derived for the minimum number of frames that need to be integrated to ensure proper tracking. This is verified using simulated data, and its importance is discussed.

In Chapt. 8 the limitations imposed by atmospheric turbulence are derived. Path-difference fluctuations are described based on the spectral method of wave propagation, and previous observations of fringe motions are presented. The visibility loss due to fringe motion is outlined in §8.1. This introduces the concept of coherent and incoherent integration of fringes, and thus the motivation for using power spectrum estimates and the Fourier transform. It is shown that pathlength motions of one wavelength will completely destroy the observed fringe, but that by recording fringe power spectra, each taken on short time scales, the fringe amplitude can be integrated. In §8.4 the limitations of active and passive fringe tracking are then presented.

Chapter 6

The Detector Response

The process of detection limits the information that can be obtained from a channelled spectrum. The spectrum becomes smoothed and sampled by the detector: it is smoothed across the width of each pixel and then sampled there. The averaging reduces the visibility of a fringe that has a high spatial frequency, the sampling may cause aliasing of the fringe frequency, and the extent of the sampled data determines what resolution is available in the spectrum.

In this chapter the relationship between the discrete and continuous functions will be examined. The spatial frequency response of the detector can be thought of in terms of a transfer function. This will be illustrated by comparing the integral Fourier transform with the Discrete Fourier Transform, in both the domain of the detector and in the spatial frequency domain. The examples that are discussed characterise the performance of an idealised array operating under conditions of high signal-to-noise. Low signal-to-noise issues will be reviewed in the following chapter.

6.1 The Transfer Function of an Array Detector

The relationship between the discrete and integral transforms is described in Appendix F, where descriptions may also be found of the Rectangle function $\Pi(x/\Delta x)$, the Sha function $\text{III}(x/\Delta x)$, and their Fourier transforms. These functions are used in the discussions which follow.

It will be assumed that the detector is one dimensional in the coordinate ξ , that the spectrum is sampled by pixels of width $\Delta\xi$, that there are a total of N pixels in the array, and that there are no gaps between pixels. The mapping between detector coordinate ξ and wavenumber κ would ideally be linear, but is not necessarily so.

The sampled function

If $h(\xi)$ represents the intensity of the spectrum incident across the detector then the response of the array, $h_\theta(\xi)$, may be expressed as

$$h_\theta(\xi) = \underbrace{\Pi\left(\frac{\xi}{N\Delta\xi}\right)}_{\text{truncated}} \underbrace{\frac{1}{\Delta\xi} \text{III}\left(\frac{\xi}{\Delta\xi}\right)}_{\text{sampled}} \underbrace{\left[\frac{1}{\Delta\xi} \Pi\left(\frac{\xi}{\Delta\xi}\right) * h(\xi)\right]}_{\text{filtered data}}.$$

The term in square brackets makes explicit the averaging of the intensity due to the finite pixel width, $\Delta\xi$, with the asterisk ‘*’ denoting a convolution. The Sha function $\text{III}(\xi/\Delta\xi)$ samples the averaged intensity at the pixel locations. The rectangle function $\Pi(\xi/N\Delta\xi)$ truncates the array so that it has a length of only N pixels.

The spatial frequency components of the sampled data

The complex spatial frequency spectrum of the sampled intensity, $H_\theta(s)$, is simply the Fourier transform of $h_\theta(\xi)$. This represents the information available from a Discrete Fourier Transform.

$$H_\theta(s) = \underbrace{N\Delta\xi \frac{\sin(\pi N s \Delta\xi)}{\pi N s \Delta\xi}}_{\text{resolution}} * \underbrace{\text{III}(s\Delta\xi)}_{\text{aliasing}} * \underbrace{\left[\frac{\sin(\pi s \Delta\xi)}{\pi s \Delta\xi} H(s)\right]}_{\text{delay envelope}} \quad (6.1)$$

where $h(\xi)$ and $H(s)$ are Fourier transform pairs. The separate terms in this equation may be interpreted as follows:

The term in square brackets represents the true spatial frequency spectrum $H(s)$ tapered by the transform of the pixel response—the width of the taper being dependent upon the width of a single pixel, $\Delta\xi$. This taper has its first null at $s = 1/\Delta\xi$. This corresponds to the delay envelope discussed in §2.3, and sets the coherence length for the observations. The convolution with the first sinc function represents the finite resolution of the array, determined by the length of the series sampled by the detector, $N\Delta\xi$. The longer the series the higher the resolution, and as the series length approaches infinity the sinc function approaches a delta function. The convolution with the Sha function yields a periodic repetition of the whole spectrum along the s axis. This is due to the sampling in the ξ domain and may give rise to aliasing.

6.1.1 Aliasing and the delay envelope of a pixel

The amount of aliasing depends both on the sampling interval $\Delta\xi$ and the highest spatial frequency contained in the incident light $h(\xi)$. However, the pixels spatially filter $h(\xi)$ and thus the relationship between this filter and the sampling period will determine the aliasing. Consider an example.

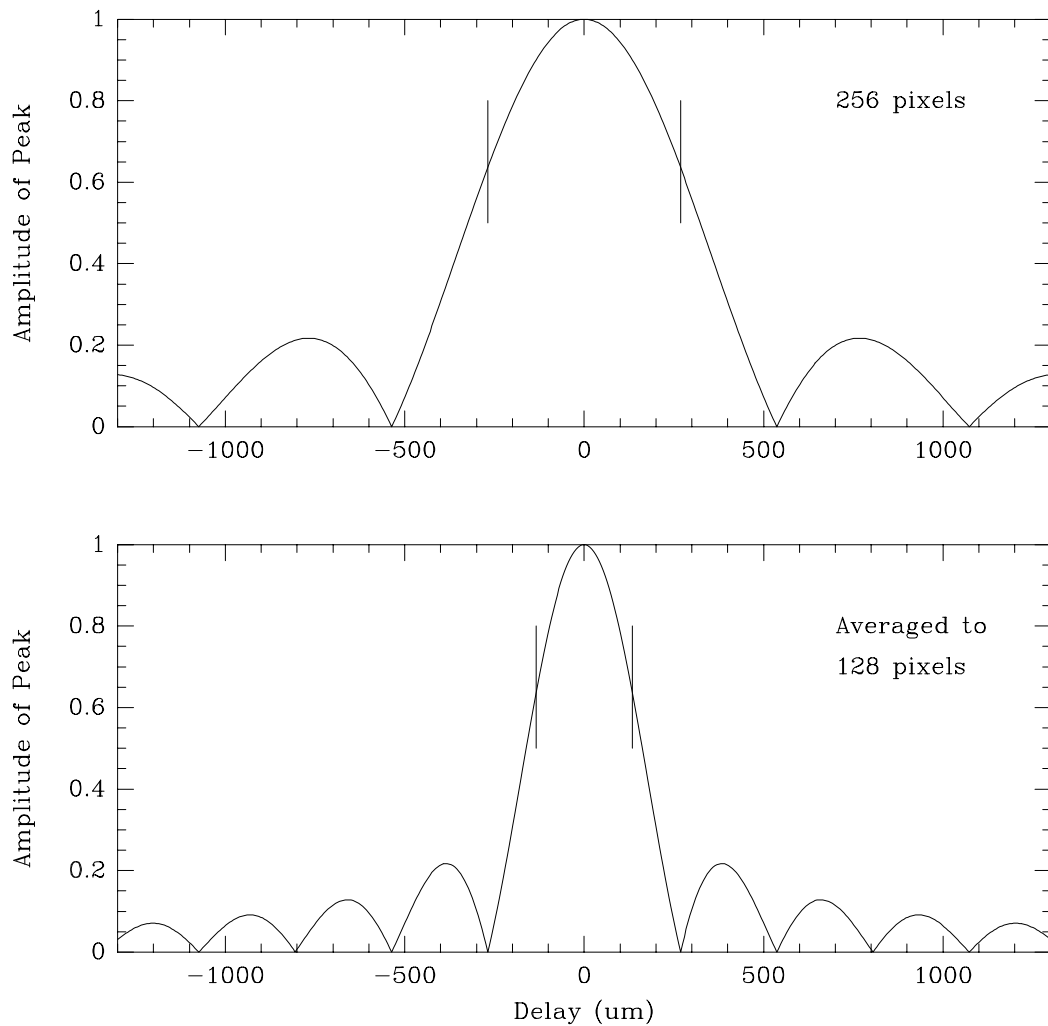


Figure 6.1: Aliasing and the delay envelope. The upper figure represents the delay curve for 256 samples evenly spaced in wavenumber from 440 nm to 550 nm. The vertical lines are halfway out to the first null and indicate the spatial frequency where aliasing would begin. In the lower figure 128 pixels are used instead, the bandwidth per pixel is larger, and the delay curve is therefore narrower.

If 256 pixels with size $\Delta\xi$ are used to sample $h(\xi)$ over a total extent of $256\Delta\xi$, then a Discrete Fourier Transform will yield 256 spatial frequency components between the frequencies $-1/(2\Delta\xi)$ and $+1/(2\Delta\xi)$. Aliasing will occur because the pixels have a rectangular, rather than a sinc function, bandpass. The relationship between the rectangular bandpass and its delay envelope is as follows:

$$\frac{1}{|\Delta\xi|} \Pi\left(\frac{\xi}{\Delta\xi}\right) \Leftrightarrow \frac{\sin \pi s \Delta\xi}{\pi s \Delta\xi}$$

and therefore the highest spatial frequency components lie halfway out to the first null of the delay curve, since the first null lies at $s = 1/\Delta\xi$. All higher frequencies will be aliased. The sinc function is illustrated in Fig. 6.1, and aliasing is clearly seen in the simulations of Fig. 6.2 where the detected spatial frequency shows a discontinuity, and then a reversal. If the pixels had a sinc function profile then their Fourier transform would be band-limited, and there would be no aliasing. However, with a rectangular profile their transform has high frequency components well beyond the Nyquist limit.

6.1.2 The resolution in delay

Each spatial frequency sample has its own ‘coherence length.’ The DFT will not resolve the delay in a channeled spectrum to better than $1/\Delta\kappa \mu\text{m}$, where $\Delta\kappa$ is the total bandwidth in wavenumber recorded by the array. The resolution is equal to the sampling interval in the spatial frequency domain, $1/\Delta\kappa$, and each sample will respond to delays that are nearby and unresolved.

For instance, if a spatial frequency of m (fringes per N pixels) corresponds to a delay of $x_m \mu\text{m}$, then delays near that value will also be integrated at the same spatial frequency. A delay of x will generate a relative contribution of

$$A_m = \frac{\sin(\pi(x_m - x)\Delta\kappa)}{\pi(x_m - x)\Delta\kappa}.$$

At $x = x_m$ the contribution is 1.0, but as x increases the contribution decreases and then falls to zero when $x = x_m + 1/\Delta\kappa$. That does not mean that the fringe signal disappears, but that the delay begins to be integrated at the next spatial frequency, $m+1$. As a rule of thumb therefore, the delay is integrable at a single spatial frequency so long as the change in delay is less than $1/\Delta\kappa$.

The spatial frequency where aliasing begins is dependent on the *ratio* of the sample spacing to pixel size. If the pixel width is much smaller than the sample spacing then the delay curve will be quite broad and the cut-off frequency would intersect the delay curve high up. When the pixel size is equal to the sample spacing the cut-off frequency lies half-way out to the first null—this is the physical layout for the array detector being considered. Finally, if the pixels have a width twice the sample spacing, so that the samples are overlapping, then the cut-off occurs at the first null in the delay curve.

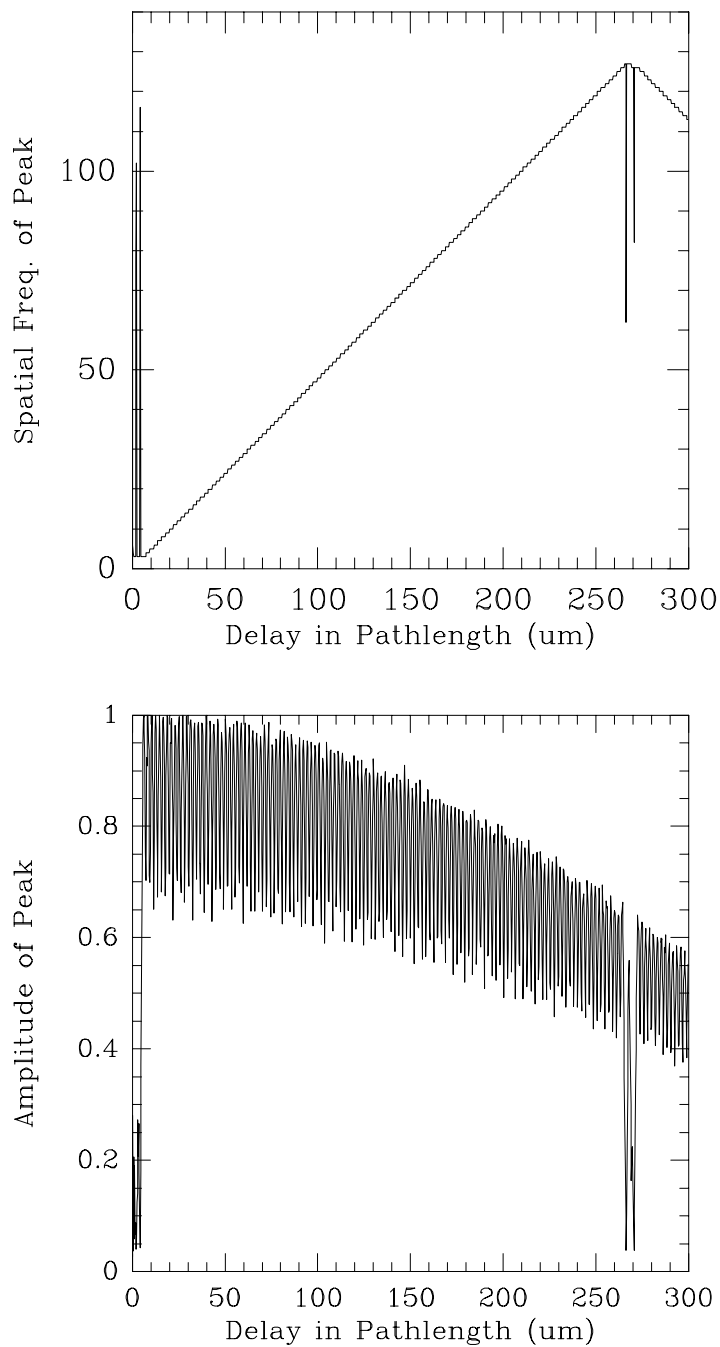


Figure 6.2: The Discrete Fourier Transform of a fringe for which the dispersion is constant across 256 pixels. The bandwidth of the spectrum is from 440 nm to 550 nm. The upper graph shows the spatial frequency of the peak as a function of delay. The lower graph shows the peak height versus delay—the noise being due to the picket-fence effect. This figure should be compared with Fig. 6.1.

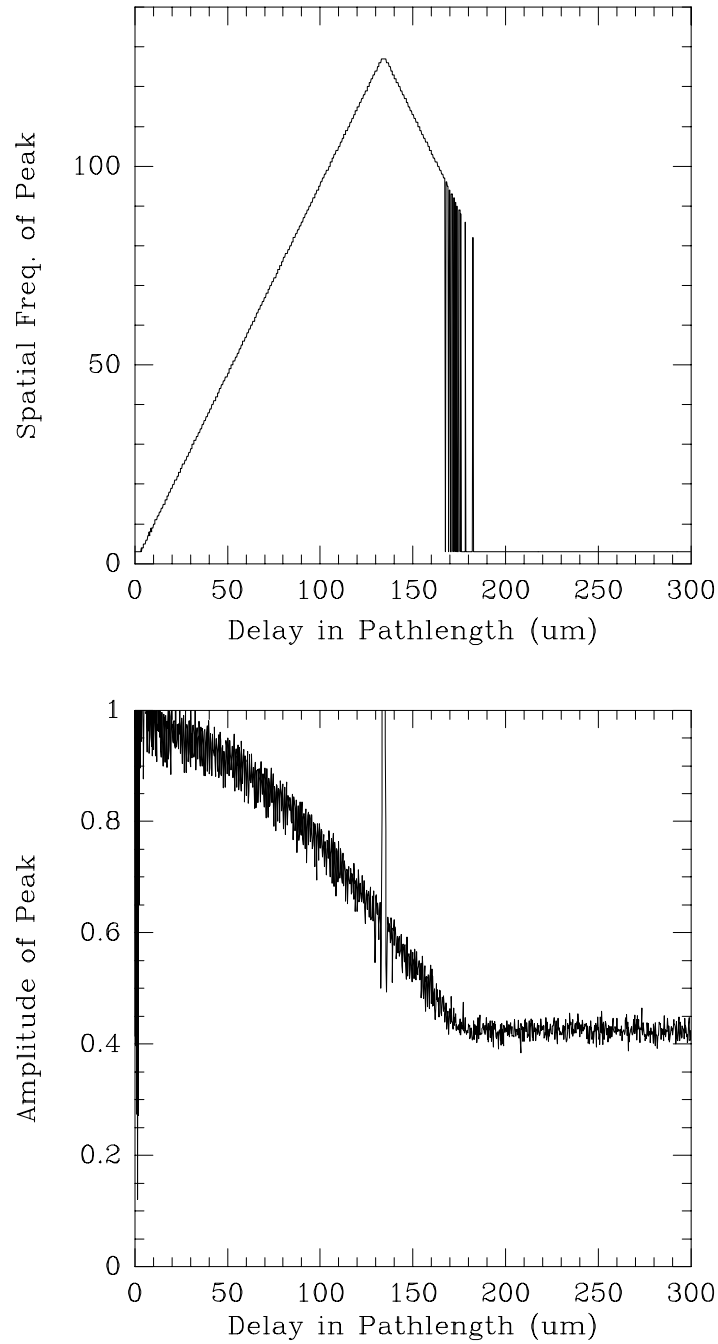


Figure 6.3: Averaging and zero padding to 128 pixels. The same data are used from the previous example, but the 256 pixels were averaged to 128 and then padded with 128 zeros. The picket-fence artifact is now reduced, but the response to higher frequencies drops off because the data have been filtered. Aliasing also is evident in these plots, obvious in the upper one where the spatial frequency of the peak is folded back to lower frequencies, and in the lower where a discontinuity in the peak height occurs. This figure should also be compared with Fig. 6.1.

6.1.3 Zero padding the array data

If the data set is artificially extended with zero valued samples, then the spacing between samples in the frequency domain is reduced. Zero padding the data does not increase the resolution in the transform, since no new information is added, but it allows the existing transform to be sampled at smaller intervals (Bergland, 1969). It also reduces the magnitude of the picket-fence effect, discussed in Appendix F and illustrated in Fig. F.1.

Averaging and padding with zeros

Consider a series of 256 samples. Rather than extending the data set with zeros, it may be averaged and *then* padded, keeping the series length constant. This method can be used with hardware implementations of the FFT where it is impossible to extend the series length, such as the Austek A41102 processor discussed in Chapt. 9. The 256 pixels can be averaged so that the total number of samples is reduced to 128; each representing two pixels and therefore a pixel width of $2\Delta\xi$. If the data is padded with 128 zeros then the series lengths comes back up to 256 samples. A 256-point transform then yields 256 spatial frequency components between the frequencies $-1/(4\Delta\xi)$ and $1/(4\Delta\xi)$. The width of the delay curve is narrowed because of the averaging, and the highest frequency component still lies halfway out to the first null (since the ratio of pixel size to sample spacing has remained constant), but now the transform is sampled at half the spacing of what it was previously.

If it were averaged further then 64 samples could be formed each representing four pixels; this could then be padded with zeros so a 256-point transform could again be used. The transform would then yield 256 samples between frequencies of $-1/(8\Delta\xi)$ and $1/(8\Delta\xi)$, with the highest frequency component still halfway out to the first null.

6.2 Visibility Loss Due to Non-Constant Dispersion

The channelled spectrum will have bright fringes wherever the optical path difference is an integer number of wavelengths, $x = m\lambda$. The *number* of fringes that appear across a fixed bandwidth is always linearly proportional to the delay, but the mapping from wavenumber to detector coordinate need not be. Using our simple model for the fringes we would write

$$I(\kappa) = I_s \left[1 + |\gamma| \cos(2\pi x\kappa + \phi_\gamma) \right] + I_b, \quad (6.2)$$

and to process this information we need to know the mapping from wavenumber to detector coordinate.

$$\kappa \longrightarrow \xi.$$

The examples in the previous section have illustrated fringes that were regularly spaced across the detector. However, most spectrometers would compress or stretch the fringe spacing. This effect is also due to the detector system, and its consequences will now be examined.

6.2.1 Linear mapping

If wavenumber is mapped linearly onto the detector coordinates,

$$\kappa = c_0 \xi$$

then it is straightforward to describe the sampled and transformed data, ignoring for the moment most DFT artifacts. If, as before, the detector has N pixels of width $\Delta\xi$, and wavelengths from λ_{min} to λ_{max} mapped onto it, then we have

$$\Delta\xi = \frac{1}{Nc_0} \left[\frac{1}{\lambda_{min}} - \frac{1}{\lambda_{max}} \right], \quad \text{or} \quad \Delta\xi = \frac{\Delta\kappa}{Nc_0},$$

where $\Delta\kappa$ is the corresponding interval in wavenumber between λ_{min} and λ_{max} .

$$\Delta\kappa = \frac{1}{\lambda_{min}} - \frac{1}{\lambda_{max}}.$$

The samples therefore lie at intervals of wavenumber given by

$$\kappa_n = \kappa_{min} + nc_0\Delta\xi, \quad n = 0, 1, \dots, N.$$

The transform determines the spatial frequency of the fringes detected across the array, that is to say m fringes per N pixels. We have therefore $x = m/\Delta\kappa$,

$$x_m = \frac{m}{Nc_0\Delta\xi}, \quad m = 0, 1, \dots, N/2.$$

and m is an index of spatial frequency. Using the expressions for κ_n and x_m we have

$$x\kappa = \frac{nm}{N} + \frac{m\kappa_{min}}{\Delta\kappa}. \quad (6.3)$$

Inserting Eq. 6.3 into 6.2 yields

$$I(n) = I_s \left[1 + |\gamma| \cos \left(\frac{2\pi m_0 n}{N} + \phi \right) \right] + I_b, \quad (6.4)$$

where

$$\phi = \frac{2\pi m \kappa_{min}}{\Delta \kappa} + \phi_\gamma.$$

The Discrete Fourier Transform of the channeled spectrum would be

$$\mathcal{I}(m) = \sum_{n=0}^{N-1} I(n) \exp \left[j \frac{2\pi m n}{N} \right],$$

whose real and imaginary parts are

$$\text{Re}[\mathcal{I}(m)] = \sum_{n=0}^{N-1} I(n) \cos \left(\frac{2\pi m n}{N} \right) = \begin{cases} N(I_s + I_b) & s = 0 \\ N(I_s |\gamma|/2) \cos \phi & m = m_0 \\ 0 & \text{else} \end{cases}$$

and

$$\text{Im}[\mathcal{I}(m)] = \sum_{n=0}^{N-1} I(n) \sin \left(\frac{2\pi m n}{N} \right) = \begin{cases} N(I_s |\gamma|/2) \sin \phi & m = m_0 \\ 0 & \text{else} \end{cases}$$

The features at $\pm m_0$ are not truly delta functions, but sinc functions whose nulls lie at the locations of the other samples in the spatial frequency domain. Examples of this transform were shown in Fig. 5.2. The peak position is linearly related to the delay, and is obvious from an inspection of the transform. Unfortunately, under most circumstances the mapping would not be linear.

6.2.2 Non-linear mapping

Most spectrometers use either a prism or a grating, neither of which have dispersions that are constant in κ . Consequently the distance between fringes will change throughout the detected spectrum, and the fringes will be partly stretched or compressed. This ‘chirp’ means that although the number of fringes would be the same, the associated frequency is more difficult to identify. The Fourier transform yields a fringe *frequency*, not a number-of-fringes. The transform is no longer a delta function—the peak is broadened and its height is reduced. For instance, if the spacing between fringes doubles from one edge of the detector to the other, then the peak would be spread between these two frequencies. This effect is illustrated in Figs. 6.4 and 6.5. In these examples the same simulation was used as in Fig. 5.2, but the dispersion is now non-constant. It is obvious that the signal has become significantly eroded. This effect is more severe the more fringes are present: at larger path differences the peak becomes progressively broader and lower in height. It becomes more difficult to detect the peak in the presence of noise, and the broadening means that the peak is less well defined. Some method of interpolation or remapping is therefore required to overcome these losses.

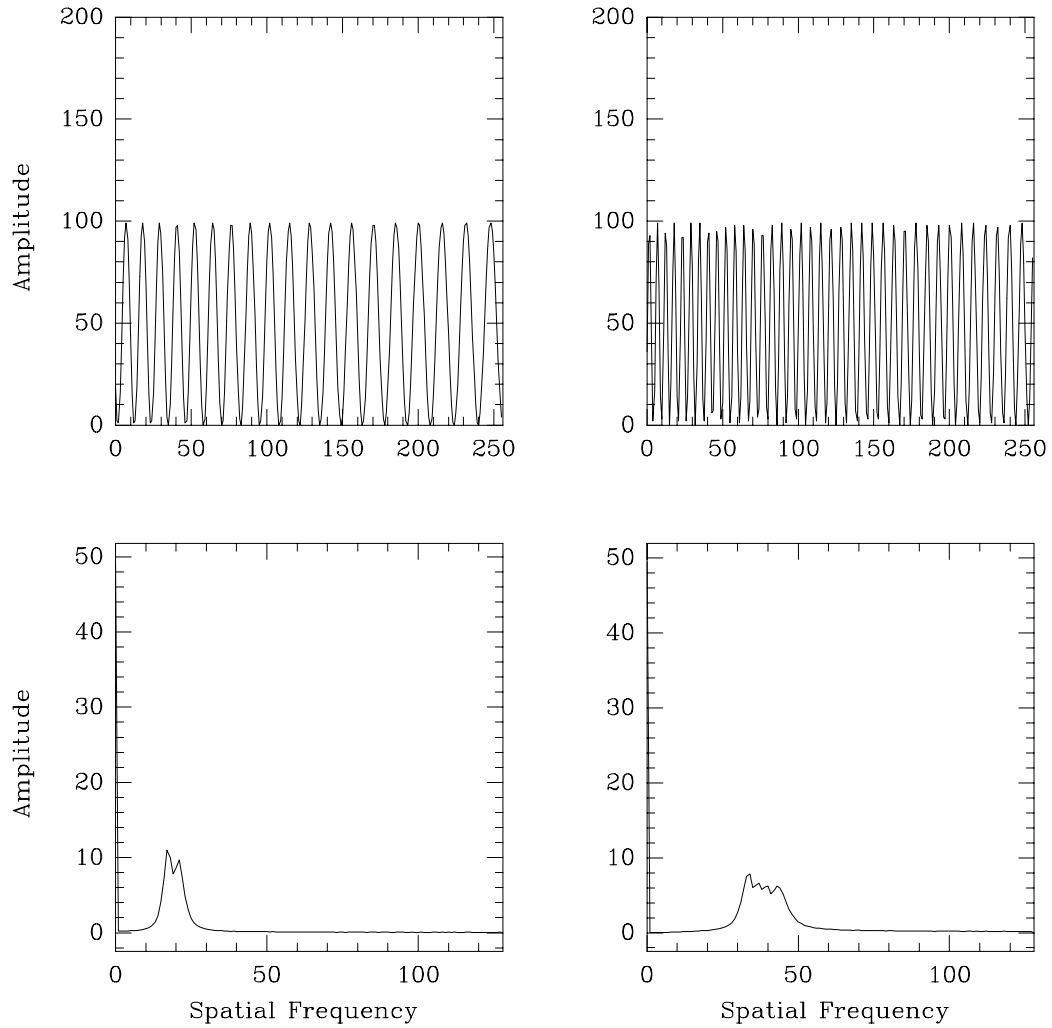


Figure 6.4: Spatial frequency response with non-constant dispersion. In this example the fringe spacing almost doubles from one side of the detector to the other. The change in fringe period across the detector causes a spread in the frequency components in its Discrete Fourier Transform. In comparison with Fig. 5.2 it is obvious that the shape of the peak has become broader and that its peak location is more difficult to identify.

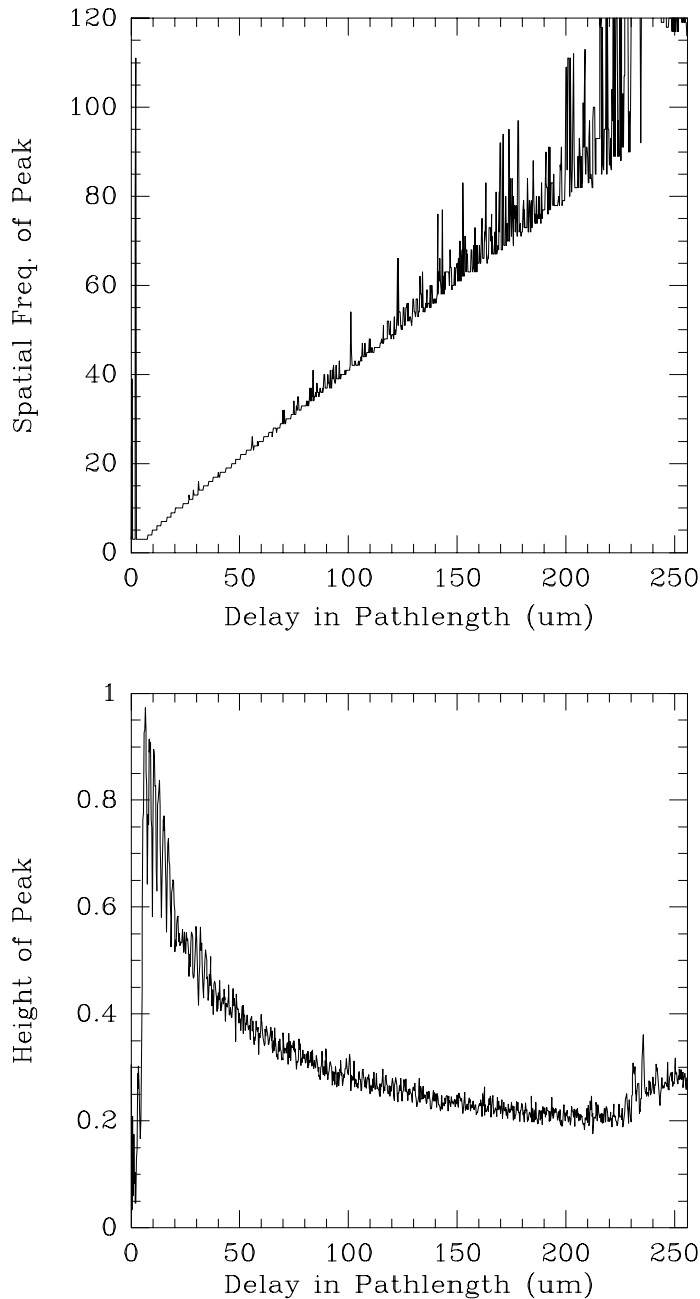


Figure 6.5: The peak height and location in the power spectrum with no re-binning. This figure should be compared with Fig. 6.2. The example is for the amplitude of the fringe, and no bias subtraction. The dispersion is that of the SF52 prism, with parameters of the spectrometer as discussed in Appendix E. The simulations model single frames of Poisson distributed data with 10000 photons-per-frame over 256 pixels, visibility of 1.0, and zero background. The signal height loses half its amplitude at a delay of about 30 μm . The increased spread in the peak can be seen from the fan-out in the graph of peak location; as the delay increases, the rms tracking error increases as well.

6.2.3 Grism spectrometer

This problem has motivated the IOTA group, at the Harvard-Smithsonian Center for Astrophysics, to design a spectrometer with a constant dispersion (Traub, 1990). It uses a combination of grating and prism, and has been called a *grism*. A prism disperses more in the blue, and a grating more in the red; thus a suitable combination of the two could yield roughly constant dispersion across the wavelength band of interest. The number of grooves per mm in the grating, the major and minor facet angles, and prism glass, must all be chosen to optimize the optical transmission in the $m = 1$ order. Although Traub lists glass and grating combinations, many of the gratings are non-standard because of his suggestions for the facet angles. It would require a separate program of research to develop such a device for the Sydney University Stellar Interferometer, and for that reason it was not considered for the current project.

6.2.4 Gridding, remapping, and interpolation

The concept of gridding data is well known in radio interferometry, where a large number of measurements are made in the Fourier plane, but at irregular intervals. For efficient processing it is necessary to re-sample the data at regular intervals. The DFT can then be used and implemented with an FFT.

The gridding is accomplished by weighing the data to improve the synthesised antenna beam, and then convolving with an appropriate function to allow sampling at the grid positions. These techniques are discussed by Thompson and Bracewell (1974), Justice et al. (1985), Thompson et al. (1986), and Perley et al. (1989). This method is computer intensive, is always performed after the data have been recorded, and never as part of a real-time servo loop.

6.2.5 Power spectra with non uniform sample spacing

It is also possible to determine power spectra from data sampled at uneven intervals, but one must then abandon the speed of the Fast Fourier Transform.

The most obvious method is to use a *Direct* Fourier Transform. This uses the samples at their uneven spacings and applies a discrete transform. The resolution, or transfer function, is then the Fourier transform of the sampling function. It may contain high sidelobes, depending on the frequency and spacing of the samples (Yen, 1956). It would require equivalent processing time to a brute force method of calculating the DFT. More efficient methods have been developed which use the FFT in non-standard ways (Press et al., 1992, §13.8). However, such methods are still computer intensive.

6.2.6 Rebinning of data

For real time processing it is simplest to rebin the data, as this requires no mathematical computations. Neither convolution nor higher order interpolation is attractive, because of the limited amount of data processing time. The rebinning re-arranges the samples, moving them to new locations with reference to a lookup table. No arithmetic is performed—the data are simply rearranged and not convolved with other data.

The look-up table

The spectrum is first split up into new pixels whose boundaries lie at even intervals in wavenumber, κ . The central wavenumber of each old pixel determines which new pixel it will be mapped to, which can be tabulated in advance and stored in a lookup table. Photon events are then rebinned upon arrival—prior to the calculation of the power spectrum.

The detector response is, of course, altered by the remapping. Stretching the data without interpolation means that some of the remapped pixels will always be empty, and compressing it means that some pixels may have twice the counts they had previously. An example of this is shown in Fig. 6.6.

Simulated remapping response

Simulations of remapped data are shown in Fig. 6.7. This represents the same simulation as in Fig. 6.5, but with the data rebinned using a lookup table. The sharp spikes and gaps in the data arise from the new detector response, which in the Fourier domain show up as noise at high spatial frequencies. The improvements from rebinning are dramatic: the peak is now sharpened and its height has increased to near maximum. This is more clearly demonstrated in Fig. 6.8 where the peak height and delay location are plotted as a function of delay.

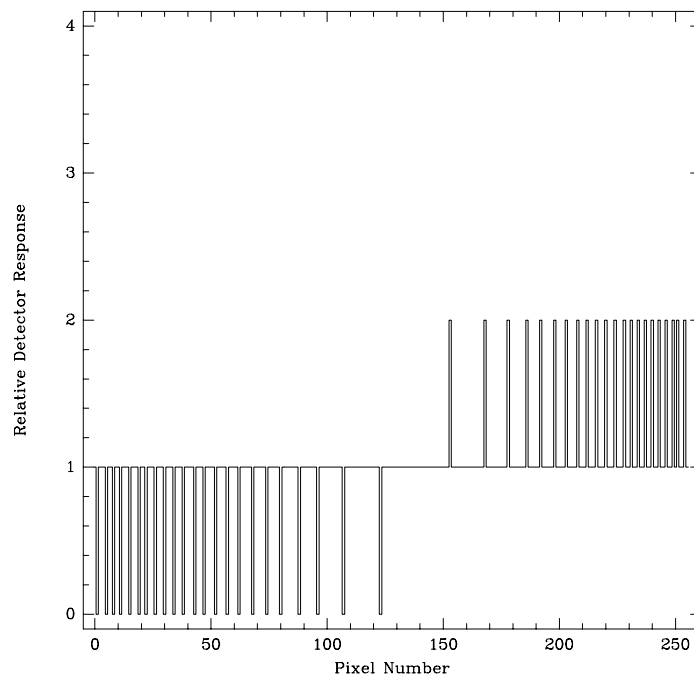


Figure 6.6: The detector response after the data have been rebinned. Here the rebinning is from 256 to 256 pixels. Notice that some of the pixels now have zero response, whereas others have twice what they had previously. In some regions the data have been stretched, and in others they have been compressed.

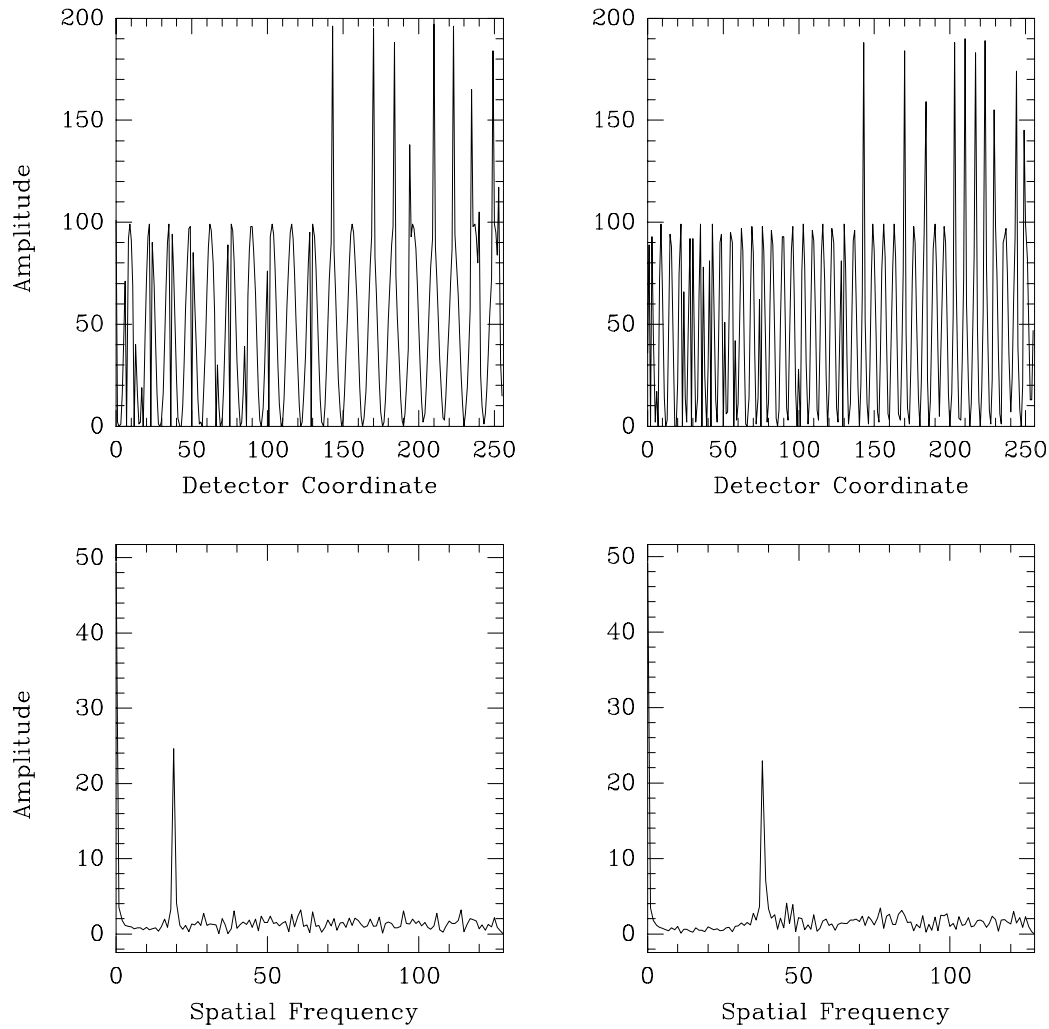


Figure 6.7: Spatial frequency spectrum after remapping for dispersion. Two sets of fringes are shown here, above, with their power spectrum, below. The fringe frequencies are the same as in Fig. 6.4, but have been remapped prior to the Discrete Fourier Transform. The improvement in the peak response is dramatic, bringing it up to half the dc level, where it would be if the dispersion had been constant in wavenumber, as in Fig. 5.2. The remapping causes some pixels to have twice the amplitude, creating sharp features in the data and giving rise to high frequency features in the spectrum.

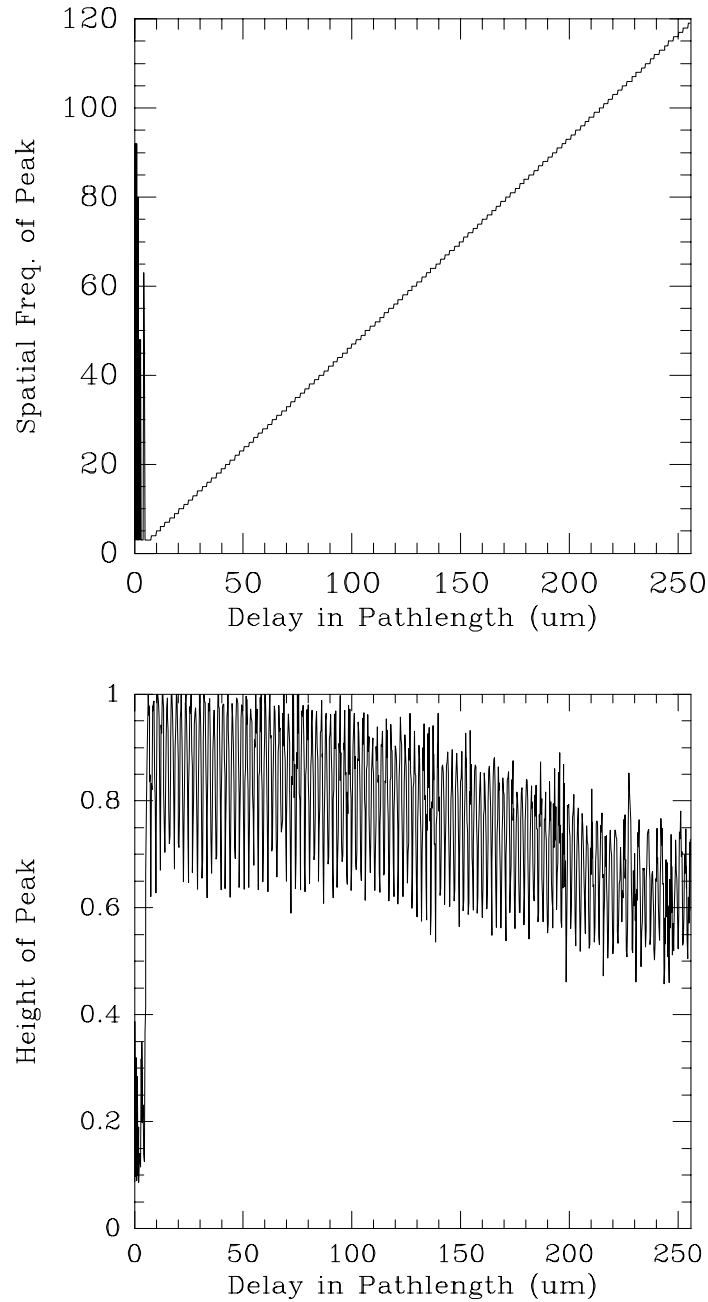


Figure 6.8: The peak height and location in the power spectrum after rebinning the data so it is mapped linear in wavenumber. The same parameters were used as in Fig. 6.5, but in this case remapping is used. The signal strength has now increased noticeably, and the ‘picket-fence effect’ is even visible in the spectrum—a clear sign that the rebinning is working well.

6.3 Conclusion

This chapter has discussed how an array detector would transform the information that could be obtained from a channeled spectrum. It was shown that the transfer function is set by the pixel width $\Delta\xi$ and the length of the array, $N\Delta\xi$. The pixel size determines the resolution in an image, and in the case of channeled spectra it determines the largest path difference that can be observed: increasing the pathlength difference brings the fringes closer together, and when the fringe spacing approaches the pixel size the fringe visibility drops to zero. If the total bandwidth across the detector is $\Delta\kappa$, where $\kappa = 1/\lambda$, and if the dispersion is constant then the resolution in delay is $1/\Delta\kappa$. A bandwidth of $\Delta\lambda=100$ nm, centered at 450 nm, therefore yields a resolution of $2.0\ \mu\text{m}$ in delay. If the length of the data set is doubled by zero padding then the resolution is not improved, but the delay can then be sampled at intervals of $1.0\ \mu\text{m}$.

In the spatial frequency domain the delay is represented by a peak at the fringe frequency. It was shown that if the dispersion of the spectrometer is non-constant then the peak will be broadened and reduced in height. This effect is more pronounced at larger delays since the fringes become spread over more and more spatial frequencies. However, it was shown that if the data is properly rebinned prior to the FFT then the response will be almost as if the dispersion had been constant.

Chapter 7

Fringe Detection at Low Light Levels

A photon counting camera detects the channeled spectrum as an array of photon counts that have been summed over a given interval. At low light levels there may be far fewer photon events than pixels, and photon-starved conditions are said to exist when there is less than one photon per spectral channel, per coherent integration time τ , per coherence region r_0 (Shao and Colavita, 1992). This is illustrated in Fig. 7.1. It is apparent that at low light levels the model of Eq. 5.1 does not adequately represent the data: the counts are discrete random events, are different in each interval, and are best described by their statistical behavior.

In this chapter the low light limitations of group delay tracking will be determined from the statistics of the photon counts. The limitations of power spectrum estimates using the DFT has been extensively studied for use in speckle interferometry and is reviewed in §7.2. The unbiased estimate of the power spectrum will be derived and the signal-to-noise ratio will be given.

The signal-to-noise ratio of the modulus of the transform, rather than its modulus squared, is reviewed in §7.3. This has been largely neglected in stellar interferometry because an unbiased estimate for the modulus does not exist. Nevertheless, it has been studied by Walkup and Goodman (1973) using Rician statistics. These statistics were originally derived for communications theory to describe the probability distribution of the envelope of a noisy sine wave (Rice, 1948), and are also used in the analysis of fringe finding in radio VLBI observations (Thompson et al. 1986). The modulus is of interest here because it is estimated by the data processing described in Chapt. 9 and is used in the experiments of Part IV of this thesis.

The probability of error in tracking is discussed in §7.4, and is the principal result of this chapter. A figure of merit is derived based on the signal-to-noise predictions, and simulations are used to verify the result. Finally, the extreme low light

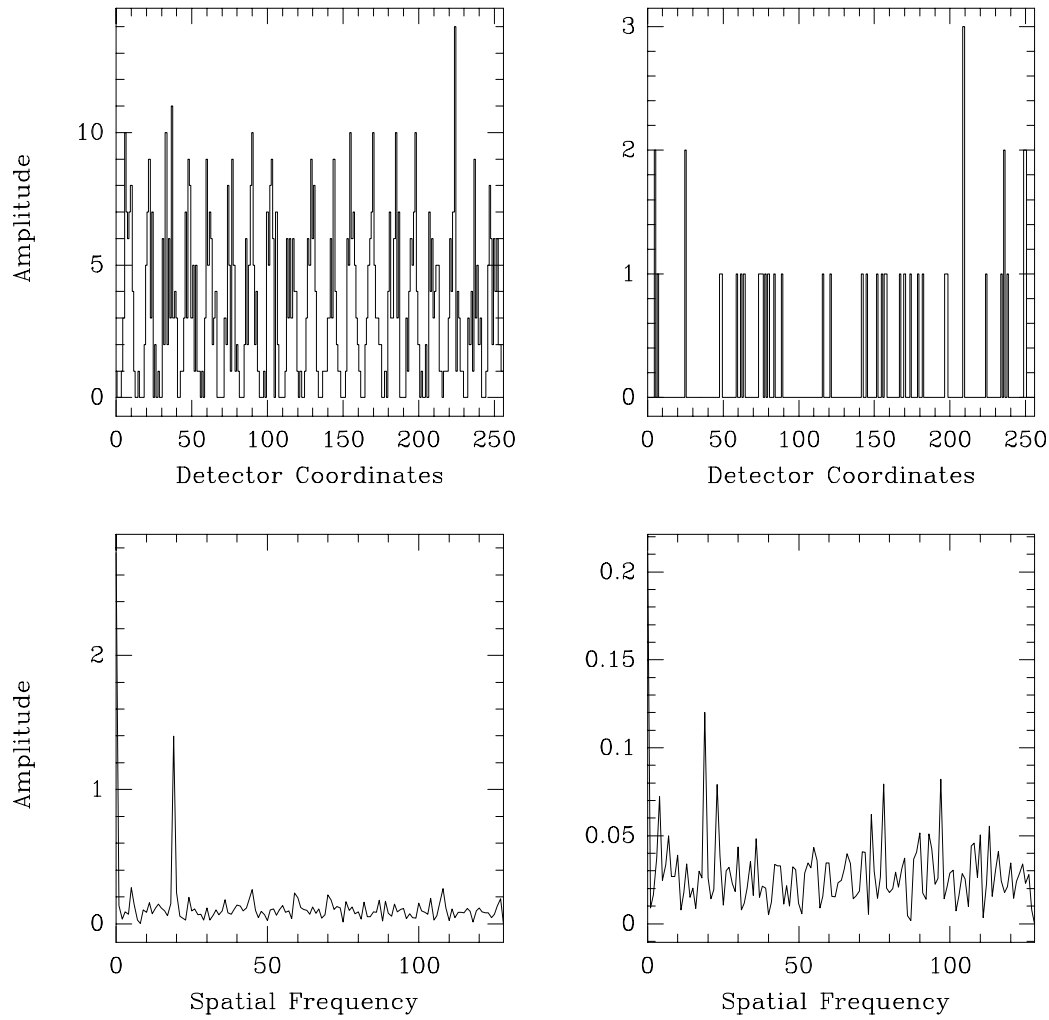


Figure 7.1: Examples of photon rich and photon starved data, showing a data set and the amplitude of its Discrete Fourier Transform. The data set on the left represents about 750 photons. The data set on the right has roughly 60 photons. The fringe frequency in each example is the same, but the noise in the photon starved case is significantly larger.

level failure of DFT estimates is described in §7.5, and the method of q -clipping is suggested to improve the performance.

The only published discussions of the signal-to-noise characteristics of group delay tracking have been those of Shao et al. (1988), and Colavita and Shao (1987). They attribute to it the same performance as that of the visibility squared estimates discussed by Tango and Twiss (1980). In their discussion each spectral channel contributes an independent estimate of $|\gamma|^2$, which is not the case for power spectrum estimates of channeled spectra that are examined in this chapter—in which all spectral channels contribute to frames within one coherence time, τ . Their interpretation of the number of photons-per-frame and what constitutes a frame is therefore different.

7.1 The Channeled Spectrum with Photon Noise

It will be assumed that the counts obey Poisson statistics, whose higher moments are listed by Tango and Twiss (1980). For high count rates the probability distribution approaches a Gaussian with the same mean and standard deviation, but for low count rates (less than 10 photons per pixel on average) the distribution is asymmetrical.

On the array detector the average counts at any element n would be proportional to the amount of light received there. We may rewrite Eq. 5.1 as follows

$$\bar{k}(n) = k_s \left[1 + |\gamma| \cos \left(\frac{2\pi n s_0}{N} + \phi \right) \right] + k_b, \quad (7.1)$$

where $k(n)$ is the photon count at the n^{th} pixel, k_s is a constant determined by the amount of light passing through the apertures, the optical efficiency of the interferometer, the area of a single array element, the integration time, and the quantum efficiency of the detector. $|\gamma|$ is the fringe visibility, s_0 is the fringe frequency, ϕ is its phase, N is the total number of elements in the detector, and k_b is the level of the background counts.

Walkup and Goodman (1973) take the Discrete Fourier Transform of Eq. 7.1 and separate it into its real and imaginary parts, $\bar{\mathcal{K}}_R(s)$ and $\bar{\mathcal{K}}_I(s)$. They then examine the statistics of each part separately. Using the fringe model of Eq. 7.1 it is easy to show that the mean values of the real and imaginary parts are

$$\bar{\mathcal{K}}_R(s) = \sum_{n=0}^{N-1} \bar{k}(n) \cos \left(\frac{2\pi n s}{N} \right) = \begin{cases} N(k_s + k_b) & s = 0 \\ N(k_s |\gamma|/2) \cos \phi & s = s_0 \\ 0 & \text{else} \end{cases}$$

$$\bar{\mathcal{K}}_I(s) = \sum_{n=0}^{N-1} \bar{k}(n) \sin \left(\frac{2\pi n s}{N} \right) = \begin{cases} N(k_s |\gamma|/2) \sin \phi & s = s_0 \\ 0 & \text{else} \end{cases}$$

The factor of 1/2 appears in the calculation because both positive and negative frequency components exist.

7.2 The DFT Power Spectrum Signal-to-Noise Ratio

Although Walkup and Goodman (1973) use the fringe amplitude as their estimator, discussed in §7.3, they continue their arguments by deriving the variance of \mathcal{K}_R and \mathcal{K}_I . This derivation closely follows that of Goodman (1984), but is more general in its scope and results in an expression for the photon bias. The signal-to-noise ratio for the power spectrum is then shown. Let us begin by noting that for Poisson deviates

$$\overline{k^2}(n) = \overline{k}(n)^2 + \overline{k}(n) \quad (7.2)$$

$$\overline{k(n)k(m)} = \overline{k}(n)\overline{k}(m), \quad m \neq n \quad (7.3)$$

and that the real part of the Fourier transform (squared) can be written as

$$\begin{aligned} \mathcal{K}_R^2(s) &= \left[\sum_{n=0}^{N-1} k(n) \cos \frac{2\pi ns}{N} \right]^2 \\ &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} k(n)k(m) \cos \frac{2\pi ns}{N} \cos \frac{2\pi ms}{N}. \end{aligned}$$

The mean squared value can then be found as follows

$$\begin{aligned} \overline{\mathcal{K}_R^2}(s) &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \overline{k(n)k(m)} \cos \frac{2\pi ns}{N} \cos \frac{2\pi ms}{N} \\ &= \left[\sum_{n=0}^{N-1} \overline{k}(n) \cos \frac{2\pi ns}{N} \right]^2 + \sum_{n=0}^{N-1} [\overline{k^2}(n) - \overline{k}(n)^2] \cos^2 \frac{2\pi ns}{N} \end{aligned} \quad (7.4)$$

where we have used Eq. 7.3 to separate terms $n = m$, and $n \neq m$. If we now insert Eq. 7.2 into 7.4 and use the identity $\cos^2 \theta = (1 + \cos 2\theta)/2$ we have

$$\begin{aligned} \overline{\mathcal{K}_R^2}(s) &= \left[\sum_{n=0}^{N-1} \overline{k}(n) \cos \frac{2\pi ns}{N} \right]^2 + \sum_{n=0}^{N-1} \overline{k}(n) \cos^2 \frac{2\pi ns}{N} \\ &= \left[\sum_{n=0}^{N-1} \overline{k}(n) \cos \frac{2\pi ns}{N} \right]^2 + \frac{1}{2} \sum_{n=0}^{N-1} \overline{k}(n) \\ &\quad + \frac{1}{2} \sum_{n=0}^{N-1} \overline{k}(n) \cos \frac{2\pi n(2s)}{N} \end{aligned} \quad (7.5)$$

The variance is therefore

$$\sigma_R^2(s) = \frac{1}{2} \sum_{n=0}^{N-1} \overline{k}(n) + \frac{1}{2} \sum_{n=0}^{N-1} \overline{k}(n) \cos \frac{2\pi n(2s)}{N}$$

In a similar manner we can derive the expression for $\overline{\mathcal{K}_I^2}(s)$ but use the identity $\sin^2 \theta = (1 - \cos 2\theta)/2$ and write

$$\begin{aligned} \overline{\mathcal{K}_I^2}(s) &= \left[\sum_{n=0}^{N-1} \overline{k}(n) \sin \frac{2\pi ns}{N} \right]^2 + \frac{1}{2} \sum_{n=0}^{N-1} \overline{k}(n) \\ &\quad - \frac{1}{2} \sum_{n=0}^{N-1} \overline{k}(n) \cos \frac{2\pi n(2s)}{N} \end{aligned} \quad (7.6)$$

with

$$\sigma_I^2(s) = \frac{1}{2} \sum_{n=0}^{N-1} \bar{k}(n) - \frac{1}{2} \sum_{n=0}^{N-1} \bar{k}(n) \cos \frac{2\pi n(2s)}{N}$$

In general the real and imaginary parts of the discrete Fourier transform have difference variances. In each case the first term in the variance is a bias dependent upon the total number of photons, and the second term samples the DFT at the second harmonic of the spatial frequency under consideration. Walkup and Goodman (1973) and Goodman (1984) show that using the fringe model of Eq. 7.1 the second term is identically zero, and therefore both \mathcal{K}_R and \mathcal{K}_I have equal variances. Their discussion then follows the thread picked up again in §7.3, but let us continue our calculations.

Photon noise and the mean value of the periodogram

Using Eqs. 7.5 and 7.6 the mean value of the periodogram estimate can be written as

$$\begin{aligned} \langle \mathcal{K}_R^2(s) + \mathcal{K}_I^2(s) \rangle &= \left[\sum_{n=0}^{N-1} \bar{k}(n) \cos \frac{2\pi ns}{N} \right]^2 + \left[\sum_{n=0}^{N-1} \bar{k}(n) \sin \frac{2\pi ns}{N} \right]^2 \\ &\quad + \sum_{n=0}^{N-1} \bar{k}(n) \end{aligned} \quad (7.7)$$

The last term is a noise bias which is independent of s and is therefore present at all spatial frequencies. It is simply the average number of photons counted across the detector,

$$\bar{N}_t = \sum_{n=0}^{N-1} \bar{k}(n).$$

It is common practice in speckle interferometry to use an unbiased estimator for the power spectrum, so that the image may be accurately reconstructed (Dainty and Greenaway, 1979; Dainty, 1984). The bias is easily predicted and can be subtracted from the mean to yield an *unbiased* estimator for the periodogram:

$$Q(s) = \langle \mathcal{K}_R^2(s) + \mathcal{K}_I^2(s) \rangle - \sum_{n=0}^{N-1} \bar{k}(n), \quad (7.8)$$

The difference between biased and unbiased power spectrum estimates is shown in the accompanying figures. Fig. 7.2 shows how the subtraction of the bias alters a single power spectrum, and Fig. 7.3 shows its effect when power spectra are summed together. The subtraction removes the noise floor, giving the noise a zero mean value, but does not alter the structure in the spectrum—the differences in height between the noise peaks and signal peak remain the same. It is interesting to note therefore that bias subtraction would serve no useful purpose in group delay tracking.

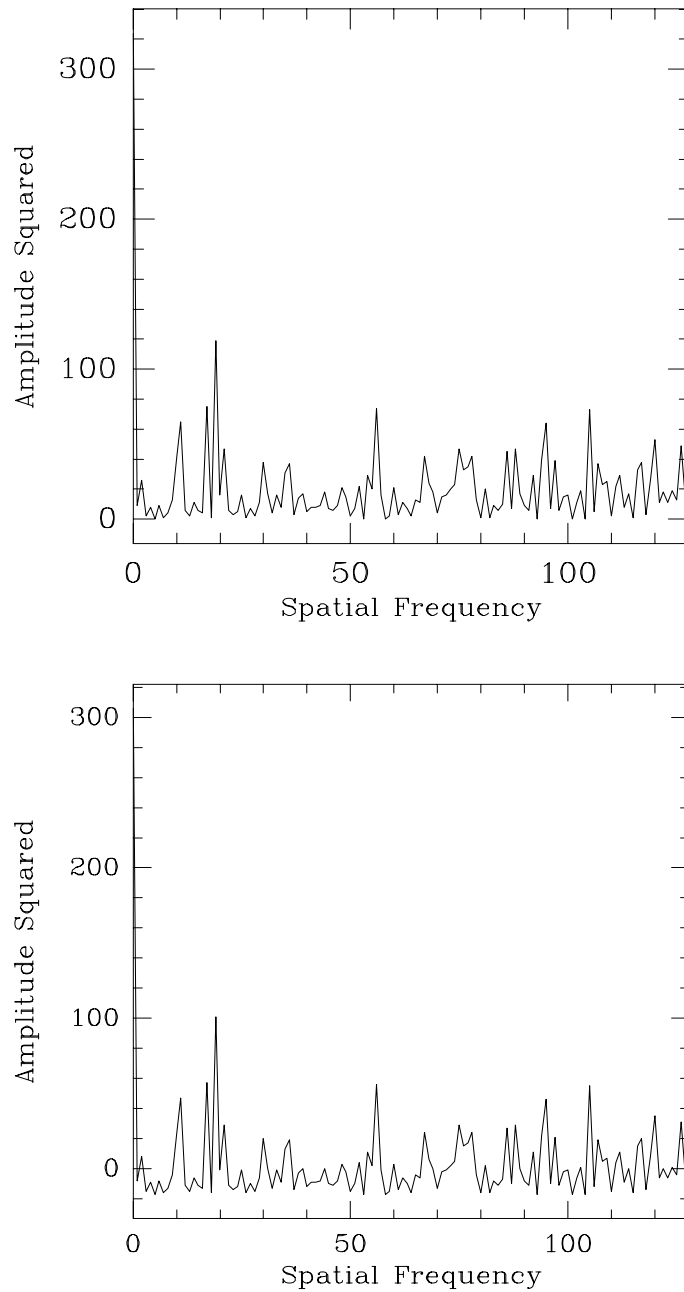


Figure 7.2: Biased and unbiased estimates of squared-amplitude. In the above examples there are 18 photons-per-frame, and one frame only. The exact same simulation is used in each case, except that the lower figure has the photon noise bias removed. If unbiased estimates are used then the noise will average to a zero mean when a large number of frames are integrated.

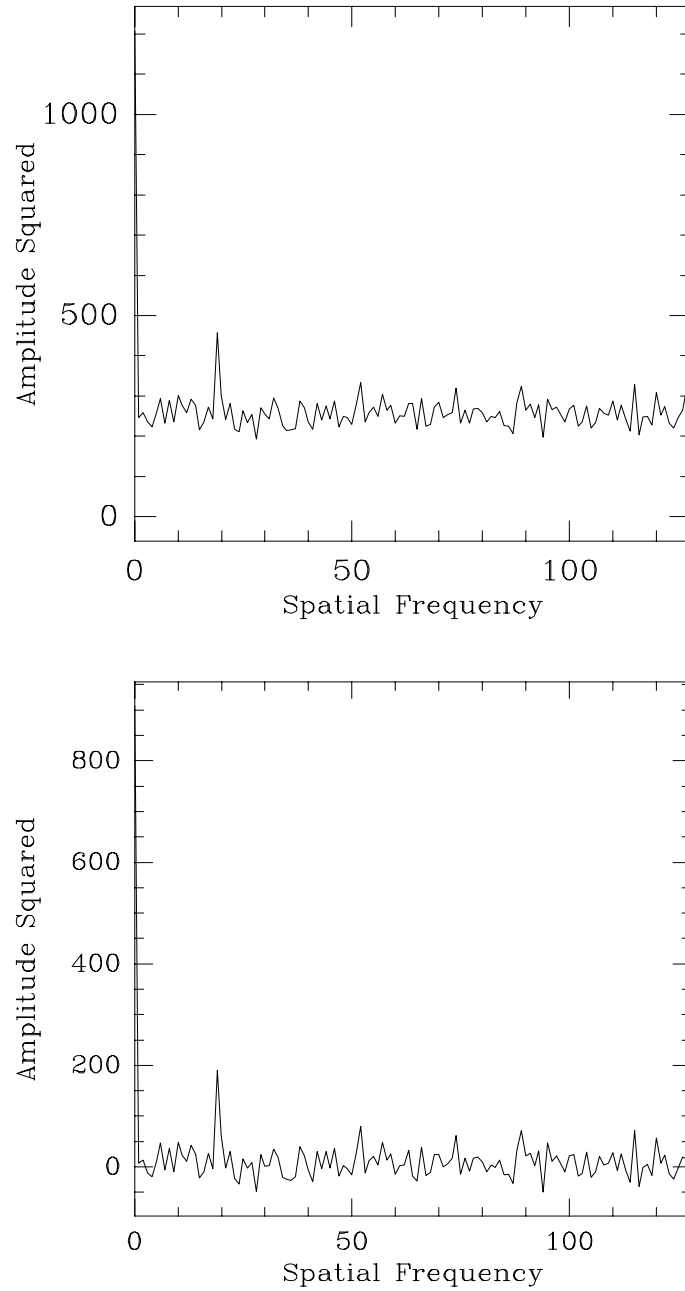


Figure 7.3: Integrated biased and unbiased estimates of squared-amplitude. Here we have simulations with an average of 3 photons-per-frame, and an integration of 100 frames. Once more the same simulations are used in each case, but the lower figure has the noise bias subtracted from each frame prior to it being summed with the others.

Higher order statistics of the periodogram

The effect of photon noise on speckle images was originally studied in a suite of papers by Goodman and Belsher (1976a, 1976b, 1977). Roddier (1986) reviews the derivation in his comparison between pupil-plane and image-plane interferometry. His paper is useful for a table which lists the terms used in Goodman and Belsher's work. A more compact and readable discussion, based on Roddier's review, is contained in a recent paper by Beletic and Goody (1992). The same signal-to-noise study is also contained in papers by Ayers, Northcott, and Dainty (1988), and Kulkarni, Prasad, and Nakajima (1991) as a preliminary to discussions of bispectrum imaging.

Beletic and Goody (1992) treat a special case of the signal-to-noise calculations performed by Roddier (1986), and their derivation will now be outlined. The notation is based on that used by Goodman and Belsher (1976).

$$\begin{aligned} D(s) &= \sum_{n=0}^{N-1} k(n) \exp \left[-j \frac{2\pi n s}{N} \right], \\ \Lambda(s) &= \sum_{n=0}^{N-1} \bar{k}(n) \exp \left[-j \frac{2\pi n s}{N} \right]. \end{aligned}$$

The samples, $k(n)$, are assumed to be so sparsely distributed that the data are an array of delta functions, each delta function locating the position of one photon event. An integral transform is then applied to mimic the DFT. The power spectrum estimate is nevertheless identical to Eq. 7.8,

$$\begin{aligned} Q(s) &= |D(s)|^2 - N_t, \\ \langle Q(s) \rangle &= |\Lambda(s)|^2. \end{aligned}$$

The signal-to-noise ratio is defined as the mean value of $Q(s)$ over its rms value

$$\text{SNR}(s) = \frac{\langle Q(s) \rangle}{\sigma[Q(s)]},$$

where we have

$$\begin{aligned} \sigma^2[Q(s)] &= \langle Q^2(s) \rangle - \langle Q(s) \rangle^2 \\ &= \langle Q^2(s) \rangle - |\Lambda(s)|^4. \end{aligned}$$

and

$$\langle Q^2(s) \rangle = \langle |D(s)|^4 \rangle - 2\langle N_t |D(s)|^2 \rangle + \langle N_t^2 \rangle$$

The most difficult part of the subsequent calculations is the evaluation of $\langle |D(s)|^4 \rangle$. The method of attacking this problem is attributed to Goodman and Belsher (1976) and is summarised by Roddier (1986).

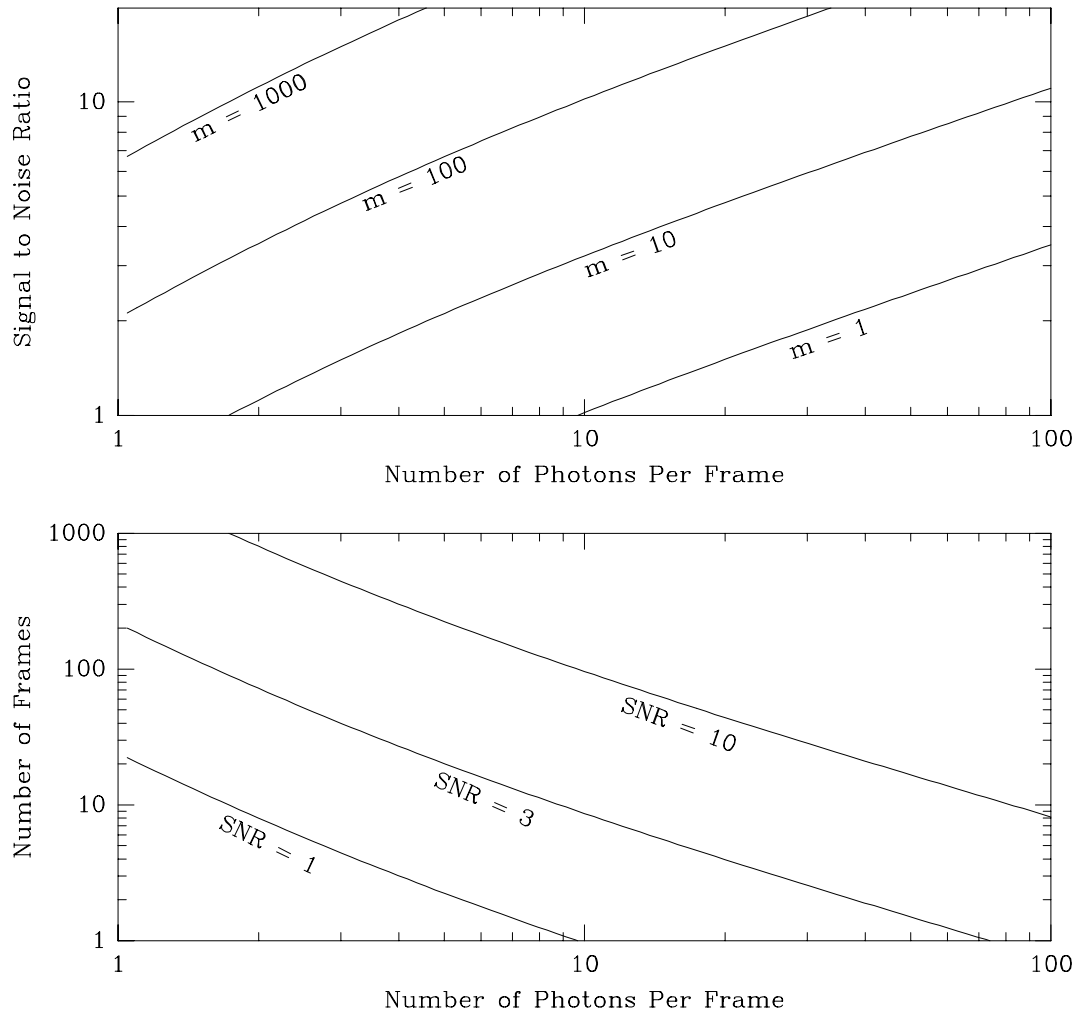


Figure 7.4: Signal-to-noise ratio for integrated frames of unbiased power spectrum estimates. It is assumed that the visibility is 1.0 and there are no background counts. The upper and lower diagrams present the same information, but expressed differently. The lower figure shows the number of frames required to maintain the signal-to-noise level constant, plotted for several different values of the SNR.

The mean value of $|D(s)|^4$

$$|D(s)|^4 = \left[\sum_{n=0}^{N-1} k(n) \exp \{-j2\pi ns/N\} \right] \times \left[\sum_{m=0}^{N-1} k(m) \exp \{j2\pi ms/N\} \right] \\ \times \left[\sum_{p=0}^{N-1} k(p) \exp \{-j2\pi ps/N\} \right] \times \left[\sum_{q=0}^{N-1} k(q) \exp \{j2\pi qs/N\} \right]$$

$$|D(s)|^4 = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} k(n) k(m) k(p) k(q) \exp \{-j2\pi[(n-m) + (p-q)]s/N\}$$

The N^4 terms in this summation can be divided up into 15 groups, evaluated in Table 1 of the appendix of Roddier (1986). The result is

$$\sigma^2[Q(s)] = N^2 + 2N|\Lambda(s)|^2 + |\Lambda(2s)|^2 \\ + 2\text{Re}[\Lambda(2s) \Lambda^*(s) \Lambda^*(s)].$$

If we ignore the presence of the half-frequency components, assuming $|\Lambda(2s)| = 0$, then we arrive at

$$\text{SNR}(s) = \frac{|\Lambda(s)|^2}{[\bar{N}_t^2 + 2\bar{N}_t |\Lambda(s)|^2]^{1/2}}$$

Now, including the fringe model and assuming there is zero background count, we set $|\Lambda| = N_t|\gamma|/2$. Furthermore, if we integrate M frames of data we get a \sqrt{M} improvement in signal-to-noise. We have then

$$\text{SNR}(s) = \frac{1}{4} M^{1/2} \frac{N_t |\gamma|^2}{\left[1 + \frac{1}{2} N_t |\gamma|^2\right]^{1/2}}, \quad \text{where} \quad M = \left(\frac{T}{\tau}\right). \quad (7.9)$$

T is the total integration time and τ is the allowed time for coherent integration of the fringes. This is the principal result of this section and is plotted in Fig. 7.4. It is, of course, the signal-to-noise ratio for delay tracking when using a DFT power spectrum estimate.

Comparison with V^2 signal-to-noise ratio

Shao et al. (1988) quote Tango and Twiss (1980) and state that for group delay tracking the signal-to-noise ratio for visibility squared estimates is

$$\text{SNR}(V^2) = \frac{1}{4} M^{1/2} \frac{N_f |\gamma|^2}{\left[1 + \frac{1}{2} N_f |\gamma|^2\right]^{1/2}}.$$

where M is defined differently than before. It is now the number of separate coherence regions where photons have been gathered, and N_f is the number of photons gathered

per region. These regions include aperture area, integration time, and the number of spectral channels.

$$M = \left(\frac{D}{r_0}\right)^2 \left(\frac{T}{\tau}\right) \left(\frac{BW}{\Delta\lambda}\right)$$

The aperture of diameter D is subdivided into r_0 sized regions that are processed separately, and the total bandwidth BW is subdivided into channels of width $\Delta\lambda$, also processed separately. The only difference between this estimate and the one derived in this section, Eq. 7.9, is the way in which information in the separate spectral channels is processed.

7.3 The DFT Modulus Signal-to-Noise Ratio

If the estimator is now the amplitude of the Discrete Fourier Transform, rather than its amplitude squared, then the statistics are somewhat different. This has been reviewed by Thompson et al. (1986, §9.3) for analysis of VLBI fringe detection, but for the moment let us follow our previous discussion following Walkup and Goodman (1973). They use as their estimator the amplitude of the DFT,

$$\hat{C} = [\mathcal{K}_R(s)^2 + \mathcal{K}_I(s)^2]^{1/2},$$

and examine it by analogy with a similar problem: the statistical properties of a sine wave plus random noise, exhaustively treated by S. O. Rice (1948). His paper served as an extension of an earlier monograph on the mathematical analysis of random noise, which he published in two parts (Rice, 1944; Rice 1945), and which has been published together in a volume of selected reprints, edited by N. Wax (1954).

The Rician distribution

In his analysis Rice looked at electrical signals of the form

$$r(t) = |\Lambda| \cos(\omega_c t + \theta) + n(t),$$

where the amplitude $|\Lambda|$ and the frequency ω_c of the carrier are known a priori. The phase θ varies randomly and the envelope is described by two quadrature components with mean amplitudes $|\Lambda| \cos \theta$ and $|\Lambda| \sin \theta$, both Gaussian distributed with identical variances of σ^2 . He showed (see N. Wax ed., 1954, pp. 236–245) that the envelope has a probability density function described by

$$p(Z) = \frac{Z}{\sigma^2} \exp \left[-\frac{1}{2\sigma^2} (Z^2 + |\Lambda|^2) \right] I_0 \left(\frac{|\Lambda| Z}{\sigma^2} \right), \quad Z > 0. \quad (7.10)$$

Where $I_0(x)$ is the modified Bessel function of order zero. This has been called the Rician distribution, or the generalized Rayleigh distribution. In the limit as $|\Lambda|$

approaches zero the two distributions are identical.

$$p(Z) = \frac{Z}{\sigma^2} \exp \left[-\frac{Z^2}{2\sigma^2} \right], \quad |\Lambda| = 0, \quad Z > 0. \quad (7.11)$$

Walkup and Goodman (1973) argued that the complex DFT may be regarded as the sum of a signal phasor having quadrature components $\bar{K}_R(s)$ and $\bar{K}_I(s)$, plus a noise phasor $\rho(s)$. If this analogy is to apply it must be assumed that the quadrature components of the noise phasor are independent, zero mean Gaussian variates, with common variance σ^2 , and for this to be true the signal-to-noise ratio must be much greater than 1. However, they state that simulations yield results that are still satisfactory for data sets with as few as 30 photons.

When the assumptions are valid the length of the noise phasor will obey a Rayleigh probability-density function ($|\Lambda| = 0$) and its phase will be uniformly distributed on $(-\pi, \pi)$. The resultant signal-plus-noise then has Rician statistics ($|\Lambda| \neq 0$) whose moments may be written

$$\langle Z^n \rangle = (2\sigma^2)^{n/2} \Gamma \left(\frac{n}{2} + 1 \right) {}_1F_1 \left(-\frac{n}{2}; 1; -\rho \right)$$

where $\rho = |\Lambda|^2/2\sigma^2$ is the signal-to-noise ratio, $\Gamma(x)$ is the Gamma function and ${}_1F_1(a; b; x)$ is a hypergeometric function given by

$${}_1F_1(a; b; x) = 1 + \frac{a}{b} \frac{x}{1!} + \frac{a(a+1)}{b(b+1)} \frac{x^2}{2!} + \frac{a(a+1)(a+2)}{b(b+1)(b+2)} \frac{x^3}{3!} + \dots$$

The mean and mean-squared values of this distribution (Rice 1948, eq. 3.14) as well as its fourth moment (Thompson, 1986, eq. 9.41) are as follows

$$\begin{aligned} \langle Z \rangle &= e^{-\rho/2} \left(\frac{\pi\sigma^2}{2} \right)^{1/2} [(1+\rho) I_0(\rho/2) + \rho I_1(\rho/2)] \\ \langle Z^2 \rangle &= |\Lambda|^2 + 2\sigma^2 \\ \langle Z^4 \rangle &= |\Lambda|^4 + 8\sigma^2 |\Lambda|^2 + 8\sigma^4 \end{aligned} \quad (7.12)$$

The moments of the Rayleigh distribution can be found by inserting $|\Lambda| = 0$ into the above equations.

It should be remembered that we derived an exact expression for the mean-squared amplitude of the periodogram, Eq. 7.7. It is the mean value $\langle |D(s)| \rangle$ that would be the primary result of this analogy. It is a biased estimate of the fringe amplitude $|\Lambda|$, and there is no obvious method of constructing an unbiased estimator, as was possible with the power spectrum.

Fringe amplitude signal-to-noise

At this stage Walkup and Goodman sidestep the problem and only consider conditions of high signal-to-noise. Their expression for fringe amplitude SNR is not the mean signal height over its rms value, but an approximation. They also choose as their estimator *twice* the amplitude of the fringe, and so get a $2^{1/2}$ improvement over what could be expected. It is now possible to write the SNR for group delay tracking which uses DFT modulus estimates. They reason that when the SNR is high the mean of the DFT and its variance can be written as

$$\begin{aligned}\langle |D(s_0)| \rangle &\simeq |\Lambda(s_0)|, \\ \text{Var} [|D(s_0)|] &= 2\sigma^2.\end{aligned}$$

The analogy with our fringe model, 7.1, leads to the following:

$$\begin{aligned}|\Lambda(s_0)| &= |\gamma| \frac{Nk_s}{2}, \\ 2\sigma^2 &= N_t = N(k_s + k_b).\end{aligned}$$

where N is the total number of pixels in the array, k_s and k_b are the average number of signal and background counts per pixel respectively, and $|\gamma|$ is the visibility of the fringe. We have therefore

$$\text{SNR} = \rho^{1/2} = \left[\frac{N k_s |\gamma|^2}{4} \left(\frac{k_s}{k_s + k_b} \right) \right]^{1/2}. \quad (7.13)$$

Their estimate compares the mean height of the signal to the mean level of its own noise bias. If you assume that $\langle Z \rangle \simeq |\Lambda|$, then with the second moment of the Rician distribution we have

$$\frac{\langle Z \rangle}{\langle Z^2 \rangle - \langle Z \rangle^2} = \frac{|\Lambda|}{\sqrt{N_t}}.$$

If there are zero background counts, $k_b = 0$, then $Nk_s = N_t$ and Eq. 7.13 reduces to

$$\text{SNR} = \rho^{1/2} = \left[\frac{N_t |\gamma|^2}{4} \right]^{1/2}$$

This is the principal result of this section, but for tracking we require another figure of merit.

7.4 The Probability of Error in Tracking

The signal-to-noise ratio does not tell the full story. What is needed is a means of predicting tracking success. A more useful parameter, calculated by Thompson et al. (1986), is the probability that one or more noise peaks will exceed the amplitude of the signal. For this calculation one needs to know the distribution of the signal and noise. A similar problem, that of the detection of images immersed in speckle noise, is treated by Dainty (1971).

The ability to locate a fringe in integrated data depends on the difference between the signal height and the mean noise level. Averaging the data over a long time will reduce rms fluctuations in each, but will not increase their separation. Let us consider visibility squared estimates. If m frames of data are averaged then the mean value remains the same, but the variance is reduced. The integrated signal, lying at $s = s_0$, can be described by the probability distribution $P_s(Z)$ which has the following features.

$$\bar{Z}_s = |\Lambda(s_0)|^2, \quad (7.14)$$

$$\sigma_s^2 = \frac{1}{m} [N_t^2 + 2N_t|\Lambda(s_0)|^2]. \quad (7.15)$$

Similarly, the distribution of noise, where $|\Lambda(s)| = 0$, can be described by a distribution $P_n(Z)$ with

$$\bar{Z}_n = 0, \quad (7.16)$$

$$\sigma_n^2 = \frac{1}{m} N_t^2. \quad (7.17)$$

The exact calculation of the probability of error for m frames would pre-suppose a knowledge of the probability density function for a single frame. The brute-force method would then derive the joint-probability distribution using an m dimensional integration, where each integral was a convolution calculated from 0 to ∞ . A more elegant method would begin by determining the characteristic function, raising that to the power m , to find the joint-characteristic function, and applying an inverse Fourier transform (Davenport and Root, 1958, §4-3). Unfortunately, neither method would be easy to apply, and there is no probability distribution we could use as a starting point.

However, in the limit of very large values of m the distribution will approach a Gaussian, making the task somewhat simpler,

$$P_x(Z) \approx \left(\frac{1}{2\pi\sigma_x^2} \right)^{1/2} \exp \left[-\frac{(Z - \bar{Z})^2}{2\sigma_x^2} \right], \quad (7.18)$$

with mean value \bar{Z} and variance σ_x^2 . Success in tracking is guaranteed if the distributions $P_s(Z)$ and $P_n(Z)$ do not overlap, which would be approximately true when

$$\bar{Z}_s - \bar{Z}_n > \beta\sigma_s + \beta\sigma_n, \quad (7.19)$$

if β is 3.0, or greater. β determines the distance from the peak of the Gaussians where the overlap occurs. If we insert Eqs. 7.14, 7.15, 7.16, and 7.17 into Eq. 7.19 we arrive at

$$|\Lambda(s_0)|^2 > \frac{2N_t\beta^2}{m} \left[1 + \frac{\sqrt{m}}{\beta} \right]$$

We can now place upper bounds on the probability of error using the fringe model we developed earlier, Eq. 7.1. If $|\Lambda(s_0)| = N_t|\gamma|/2$, so that there are zero background counts, then we have

$$N_t > \frac{8\beta^2}{m|\gamma|^2} \left[1 + \frac{\sqrt{m}}{\beta} \right]. \quad (7.20)$$

If N_t is plotted as a function of β and m we can arrive at contours for the probability of tracking loss. The contours for $\beta = 3.0$ are shown in Fig. 7.5. These lines can be thought of as 3σ levels after which tracking is ensured.

An approximate solution

The delay will be incorrectly identified if the signal level drops to a height comparable with the noise. The probability that at least one of $b - 1$ noise peaks is higher than the signal is given by

$$p_e = 1 - \int_0^\infty P_s(Z) \left[\int_0^Z P_n(z) dz \right]^{b-1} dZ. \quad (7.21)$$

This integration is identical to that used by Thompson et al. (1986, Eq. 9.60). The term in square brackets represents the cumulative probability that the noise is lower than a height Z . Z can be thought of as a threshold for success or failure in a binomial distribution. If we then consider $b - 1$ samples of noise (the other spatial frequency channels) then the probability of at least one sample exceeding the threshold Z_0 is the cumulative probability raised to the $b - 1$ power. The integral over $P_s(Z)$ then represents the probability that the signal is larger than all the noise peaks, and Eq. 7.21 therefore represents the probability of tracking failure.

If the signal and noise are represented by Gaussian distributions, then it is possible to evaluate the integral. The predictions will only be useful when a large number of frames are integrated together. The integral was therefore evaluated using the distribution of Eqs. 7.18, with Eqs. 7.14 and 7.15 for the signal, and Eqs. 7.16 and 7.17 for the noise. The results are shown plotted in Figs. 7.6 and 7.7, indicated by the dotted lines.

Simulations were also used to predict the onset of tracking failure. These used the fringe model of Eq. 7.1 with Poisson distributed noise. The squared-amplitude in the power spectrum was derived, and added to other simulations until the required number of sums, m , were performed. After the integration the power spectrum was

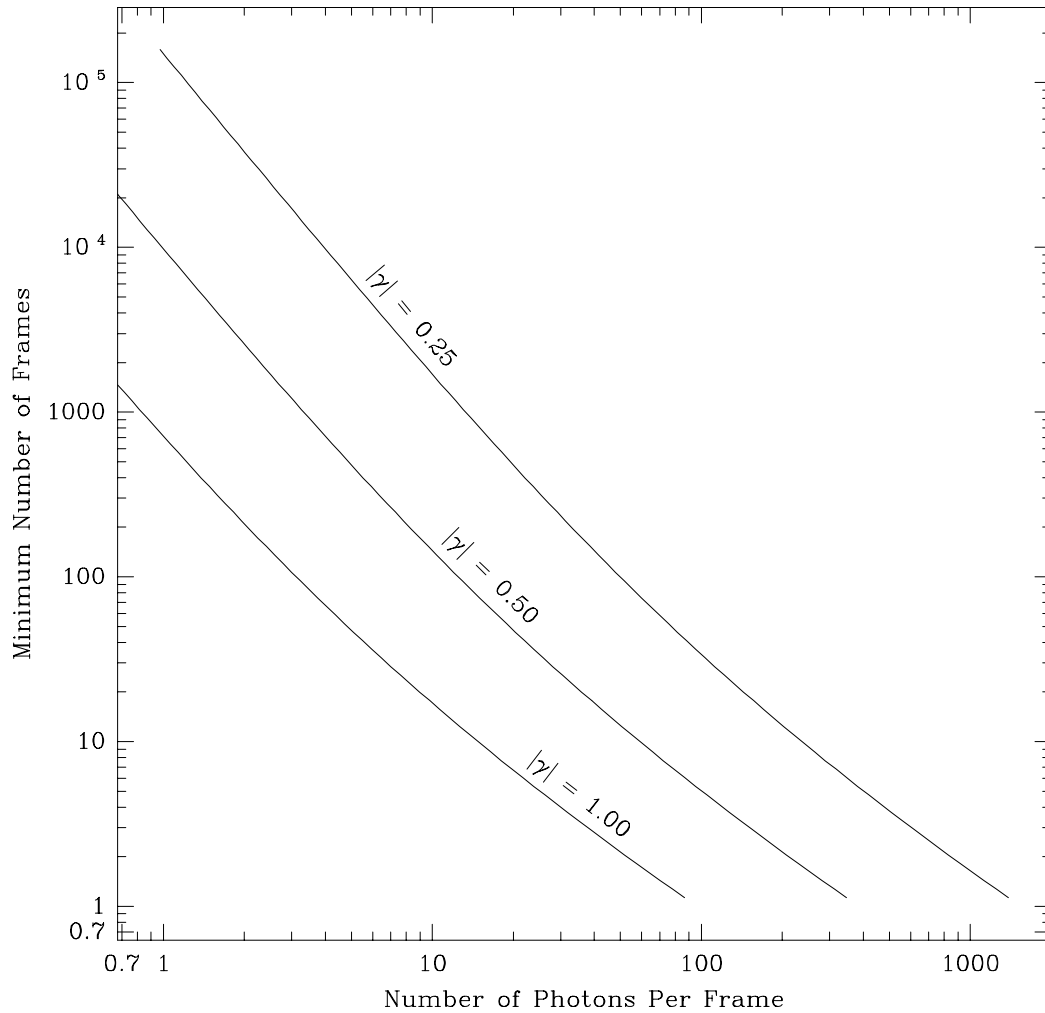


Figure 7.5: Contours for the probability of tracking loss. This models the bias corrected amplitude-squared estimates, and indicates the minimum number of frames necessary to track for a given signal level. The photons-per-frame is the number of photons gathered across 256 pixels during the coherence time of the atmosphere. This figure indicates where the 3σ levels of those distributions would intersect, as a function of visibility.

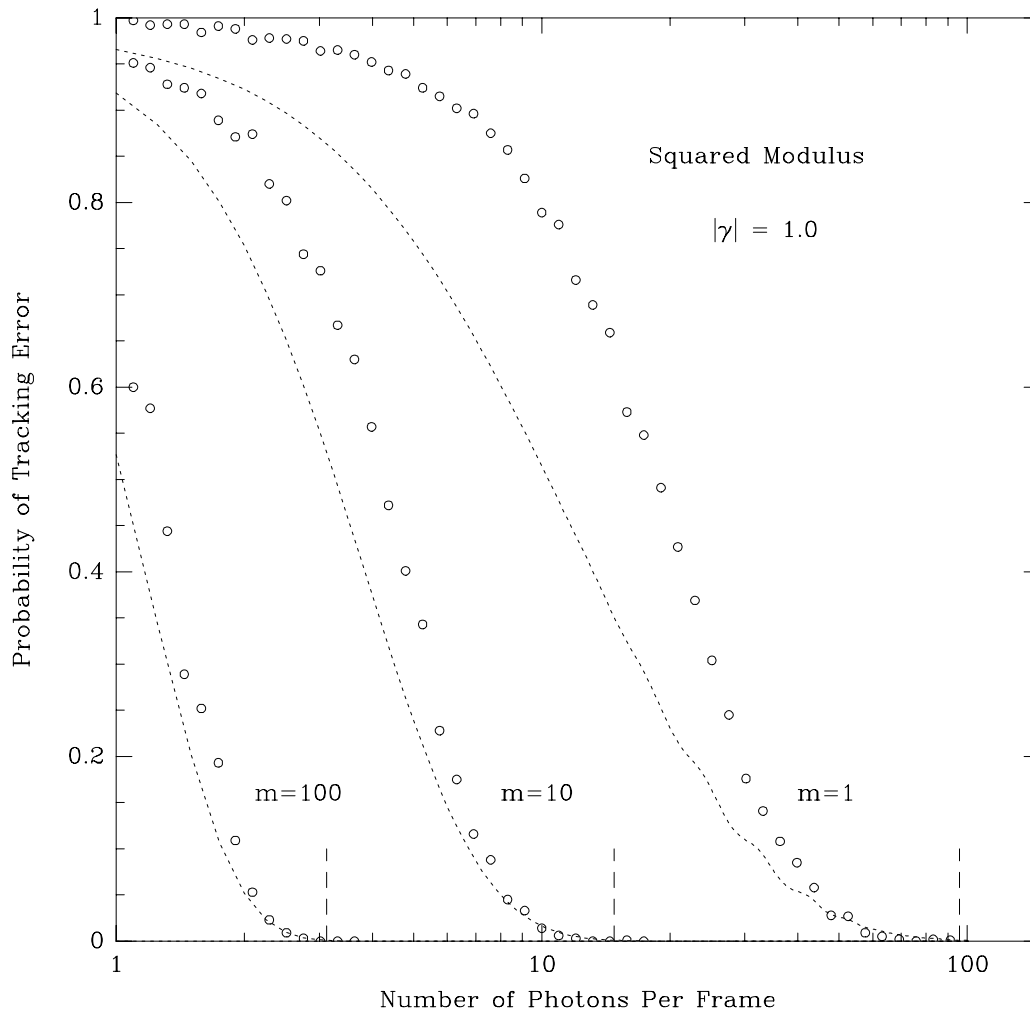


Figure 7.6: The probability of tracking error for $|\gamma| = 1.0$, as discussed in §7.4. Displayed here are the results of simulations (open circles), the approximate solution to the probability of error (dotted line), and the 3σ thresholds for tracking failure (short dashed lines, bottom). The probability of tracking error is estimated from the number of times the simulations failed to find the correct delay out of a total of 1000 trials. In these simulations the visibility used was 1.0, and bias-subtracted visibility-squared spectra are used, based on Poisson distributed data. It is apparent that the 3σ thresholds accurately predict the onset of tracking failure, and that the approximate model more closely follows the simulations only for large integrations.

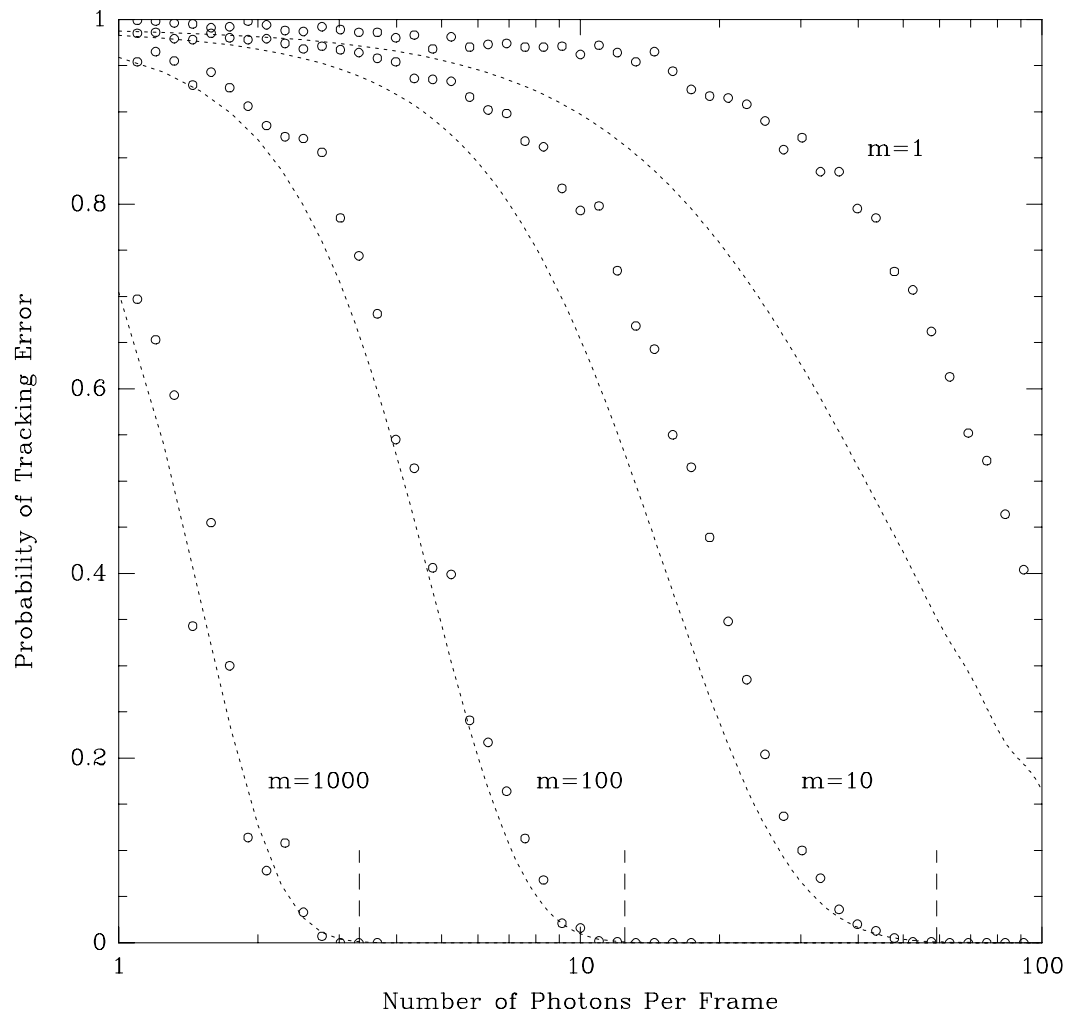


Figure 7.7: The probability of tracking error for $|\gamma| = 0.5$. The simulations and predictions shown here are identical to those shown in the previous figure, with even longer integrations included. Again, the 3σ thresholds accurately predict where tracking failure will commence.

searched to find the peak. If the peak was located correctly then the trial was recorded as a success. 1000 trials were used at each point, over which the total number of successes were counted.

It can be seen that the onset of tracking failure is the same, whether predicted by the simulations, the approximate Gaussian model, or the 3σ thresholds. The agreement between the Gaussian model and the simulations is good with large integrations, but poor otherwise.

7.5 Extreme Low light level failure of the DFT approach

It should be noted that in the realm of extremely low light levels, where there is at most one photon per frame, the discrete Fourier transform method fails completely. This is perhaps obvious, but it has some interesting implications and is worth illustrating. If the sequence to be transformed is as follows

$$f_n = \begin{cases} 1 & n = n_0 \\ 0 & \text{otherwise} \end{cases}$$

then its discrete Fourier transform and its squared modulus are

$$F_k = \exp\left(j\frac{2\pi n_0 k}{N}\right), \quad \text{and} \quad |F_k|^2 = 1, \quad 0 \leq k \leq N-1.$$

F_k is a complex quantity containing two sinusoidal terms, one real and one imaginary, whose frequencies are dependent upon the location n_0 of the sample, but whose power spectrum is *entirely* featureless with a constant height across all frequencies—with no useful information. Integrating frames of data under these conditions cannot extract delay information.

This does not imply that it is impossible to integrate frames where the mean rate is equal to (or even less than) one photon per frame. Under such circumstances it is still likely that two or more photons may occasionally be recorded in a frame, allowing information to be gathered. It is with this in mind that Dainty and Greenaway (1979) suggest removing data sets, or q -clipping the data, to improve the signal-to-noise ratio. They suggest that gains can be had by throwing out sets with zero or one photon per frame. Frames with zero or one photon reduce the signal by averaging, but contribute no information.

Removing sets with only one photon will also reduce the roundoff error in the calculation. The power spectrum for one photon event will be 1.0 ± 1 bit. For long integrations under low light conditions the roundoff errors will accumulate, and so by throwing out useless frames the errors can be reduced.

7.6 Conclusion

In this chapter there were two estimators of group delay tracking that were examined. The first in §7.2 calculated a DFT power spectrum from the channeled spectrum, and the second in §7.3 looked at the DFT modulus and therefore fringe amplitude estimates. The computer simulations of §7.4 were carried out for both methods, but yielded identical results; only the trials for power spectrum estimates were displayed in that section. The probability of tracking error was therefore identical with each of the two methods, even though the equations for signal-to-noise ratio were different.

The probability of tracking loss was discussed in §7.4. It was shown to be more informative than signal-to-noise ratios and to provide a better understanding of the limitations of delay tracking. With this in mind, it should be noted that even if the probability of tracking loss is high, perhaps 90%, it may still be possible to locate the fringe peak by a histogram of peak locations. Tracking with light levels as low as 3 photons-per-frame may be possible, but if there are at most 1 photon-per-frame then the DFT was shown to be incapable of extracting information from a channeled spectrum.

Chapter 8

Delay Tracking and Atmospheric Turbulence

In the previous chapter it was assumed that the channeled fringes were motionless when they were detected. This would only be true for short observation times. Changes in the index of refraction of dry air in the atmosphere will cause random delays in the propagation of starlight to different locations on the ground. The phase-difference changes that occur between two separated points can be directly related to motions of the white-light fringe in a Michelson interferometer. The fringe phase will be stable for timescales between 1 and 10 ms, but will change by many wavelengths during the course of an observation.

In this chapter the effects of atmospheric turbulence will be discussed and the power spectra of phase-difference changes will be described. Appendix G reviews the theory of wave propagation in a turbulent medium using the spectral representation of random variables, and provides further background for the discussion. The atmosphere's influence on active and passive delay tracking will be examined, and the results of the previous chapter will be extended to account for fringe motions.

There are several extensions to the theory which will not be treated in detail. Notably, it will be assumed when dealing with phase-difference measurements that the fields are sampled at two separated points. Interesting work has been carried out in the analysis of phase power spectra for fields sampled in large apertures. This includes the papers by Noll (1976), Hogge and Butts (1976), Greenwood and Fried (1976), Greenwood (1977), and others. However, because it will be assumed in this discussion that the apertures have a zero size, those papers will not be discussed. In effect, with point-like apertures there are no aberrations. Hogge and Butts (1976) suggest that averaging across an aperture reduces the high frequency amplitudes in the power spectra, and so assuming a zero aperture size will slightly overestimate the high frequency terms.

8.1 Visibility Loss Due to Fringe Motion

If pathlength changes occur during a time when the fringes are being recorded, then the visibility of the fringes will appear reduced. Consequently the time for coherent integration of the fringes is limited.

8.1.1 Coherent fringe detection

The recorded data consist of separate realizations, or frames, integrated over a sampling time τ wherein the atmospheric turbulence is approximately ‘frozen’. From frame to frame the optical path difference will change slowly and cause the fringes to move across the detector. If τ is too large these motions will blur the detected fringe as it is being recorded—a path change of one wavelength will cause the fringe to disappear completely. This has been discussed by Koechlin (1985) and may be illustrated as follows.

The fringe pattern will have the following form, where x is a pathlength offset from the white light fringe position, $\kappa = 1/\lambda$, I_0 is the brightness of the spectrum, and I_b is a background intensity.

$$I(\kappa) = I_s \left[1 + |\gamma| \cos(2\pi\kappa x - \alpha) \right] + I_b, \quad (8.1)$$

If x is a function of time and changes at a uniform velocity we would have $x = \bar{x} + vt$. Let us integrate the fringe pattern from a $t = -\tau/2$ to $t = \tau/2$.

$$\frac{1}{\tau} \int_{-\tau/2}^{\tau/2} I(\kappa) dt = I_s \left[1 + \frac{1}{\tau} |\gamma| \int_{-\tau/2}^{\tau/2} \cos(2\pi\kappa\bar{x} + 2\pi\kappa vt - \alpha) dt \right] + I_b$$

By expanding the cosine term, performing the integration with respect to t , and making the substitution $\Delta x = v\tau$, we arrive at

$$\frac{1}{\tau} \int_{-\tau/2}^{\tau/2} I(\kappa) dt = I_s \left[1 + \eta |\gamma| \cos(2\pi\kappa\bar{x} - \alpha) \right] + I_b,$$

where

$$\eta = \frac{\sin(\pi\kappa\Delta x)}{\pi\kappa\Delta x}. \quad (8.2)$$

If we keep in mind that $\kappa = 1/\lambda$, it can be seen that the fringe motion reduces the visibility of the fringes at shorter wavelengths more than at longer wavelengths. However, the predominant effect is to lower the visibility across the whole observed bandwidth. The sinc function is very broad and, over a limited bandwidth, we may approximate it as being constant at $\kappa = \bar{\kappa}$,

$$\frac{\sin \pi\kappa\Delta x}{\pi\kappa\Delta x} \simeq \frac{\sin \pi\bar{\kappa}\Delta x}{\pi\bar{\kappa}\Delta x}.$$

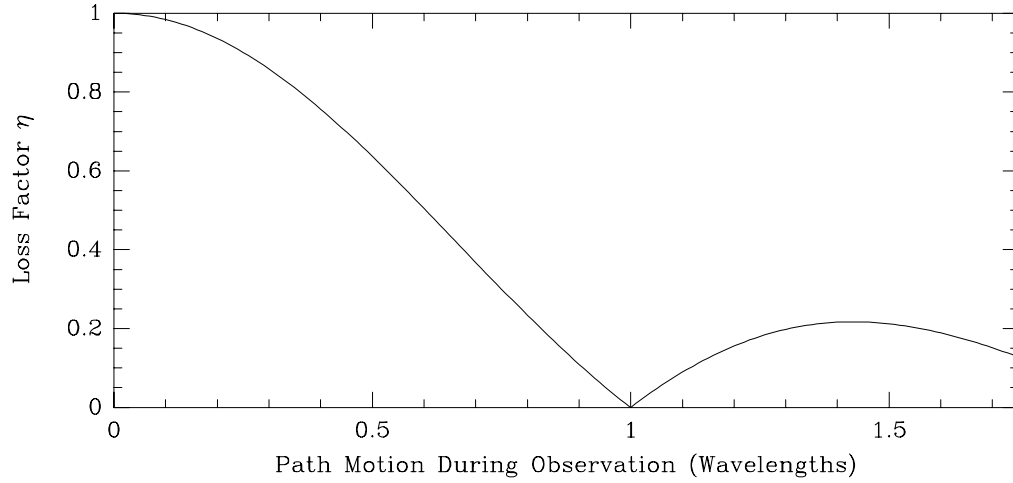


Figure 8.1: Visibility loss due to fringe motion during coherent integration. If the fringe moves during a time it is being recorded then the apparent visibility will be reduced. The loss in visibility is shown here as a function of pathlength change measured in wavelengths, $\Delta x/\lambda$.

This is plotted in Fig. 8.1. Therefore, in the Fourier domain one would expect the peak representing the fringe to be reduced by the factor η and to be reduced to zero when $\Delta x = \lambda$. Coherent integration is therefore limited to times τ where the pathlength x is stable to within a small fraction of a wavelength.

8.1.2 Incoherent fringe integration

By contrast, if τ satisfies the above restrictions then the power spectrum of the fringes will remain essentially unchanged from one frame to the next. The power spectrum will contain a peak at the spatial frequency of the fringe, and changes will not be detectable unless they become resolved. Since the resolution is only dependent on the total length of the data set, in this case the bandwidth of the channeled spectrum, the spatial frequency components can be integrated over many intervals of τ . If $\Delta\kappa$ is the bandwidth in wavenumber

$$\Delta\kappa = \frac{1}{\lambda_{min}} - \frac{1}{\lambda_{max}}$$

then the resolution in pathlength Δx available from a Fourier transform is $\Delta x = 1/\Delta\kappa$ (see the discussion concerning the rectangle function in Appendix F). In effect the Fourier transform counts chromatic fringes across the detector, but is not sensitive to fractions-of-fringes.

A bandwidth from 400 nm to 500 nm therefore represents a difference in wavenumber of $0.5 \mu\text{m}^{-1}$ and a resolution in delay of $2.0 \mu\text{m}$. The delay can therefore

change by 4 wavelengths and still be integrated at the same spatial frequency. For this reason power spectrum estimates are useful under low light conditions. When there is insufficient information during τ to establish fringe phase, it is still possible to determine the group delay.

8.2 Temporal Power Spectrum of Phase-Difference

The delay changes that are induced by the atmosphere can be characterised by a power spectrum of phase fluctuations, $W_s(f)$, whose form is dependent on the assumptions used to model atmospheric turbulence.

Tatarskii (1971, §52) discusses the spatial phase-difference spectrum which would be observed between two points separated by a distance ρ . These points can be thought of as the two apertures of a Michelson stellar interferometer. In his derivation he assumes that only small perturbations are induced in a wavefront, so that when the wave passes a point located by the vector \mathbf{r} the change in its phase can be described by $S_1(\mathbf{r})$. He writes that the time-dependent phase-difference can be described as follows:

$$\delta_\rho S(t) = S_1(\mathbf{r}, t) - S_1(\mathbf{r} + \boldsymbol{\rho}, t),$$

It is also assumed that the turbulence is ‘frozen’ in the atmosphere and is swept past the points of observation by a wind of velocity \mathbf{v}_\perp that is perpendicular to the path of the propagating wave. The phase-difference can therefore be written

$$\delta_\rho S(t + \tau) = S_1(\mathbf{r} - \mathbf{v}_\perp \tau, t) - S_1(\mathbf{r} + \boldsymbol{\rho} - \mathbf{v}_\perp \tau, t).$$

The phase structure function is then defined by

$$D_s(\rho) = D_s(|\boldsymbol{\rho}|) = \langle [S_1(\mathbf{r}) - S_1(\mathbf{r} + \boldsymbol{\rho})]^2 \rangle, \quad (8.3)$$

$$\text{or} \quad D_s(\rho) = 2[B_s(0) - B_s(\rho)].$$

where $B_s(\rho)$ is the covariance of phase described in Appendix G. The phase-difference structure function is then

$$\langle \delta_\rho S(t) \delta_\rho S(t + \tau) \rangle = \frac{1}{2} [D_s(\boldsymbol{\rho} - \mathbf{v}_\perp \tau) + D_s(\boldsymbol{\rho} + \mathbf{v}_\perp \tau) - 2D_s(\mathbf{v}_\perp \tau)].$$

Tatarski then shows that if the separation $\boldsymbol{\rho}$ and the wind velocity \mathbf{v}_\perp are parallel then the power spectrum of phase-difference fluctuations is

$$W_{\delta s}(f) = 2 \left[1 - \cos \frac{2\pi \rho f}{v_\perp} \right] W_s(f) = 4 \sin^2 \left(\frac{\pi \rho f}{v_\perp} \right) W_s(f), \quad (8.4)$$

where $W_s(f)$ is the temporal power spectrum of phase fluctuations observed at a single point. The behavior of the power spectrum is characterised by its response at low and high frequencies where the following approximations are made.

Low Frequencies

At very low frequencies f we may replace the sine squared term by its argument squared and obtain

$$W_{\delta s}(f) \simeq \left(\frac{4\pi^2 \rho^2}{v_{\perp}^2} \right) f^2 W_s(f).$$

High Frequencies

If the separation ρ is large then the fluctuations at each point are uncorrelated, and we can replace the sine squared term by its mean value of 1/2 (Tatarskii, 1971, §52, Eq. 42; Tango and Twiss, 1980).

$$W_{\delta s}(f) \simeq 2W_s(f). \quad (8.5)$$

8.2.1 Kolmogorov Turbulence

An expression for $W_s(f)$ may be derived using a modified Kolmogorov power spectrum of index of refraction fluctuations. The derivation is summarised in Appendix G and yields

$$W_s(f) = 3.28 \cdot 10^{-2} C_n^2 k^2 L v_{\perp}^{5/3} f^{-8/3}, \quad (8.6)$$

where $k = 2\pi/\lambda$ is the wavenumber of the propagating field, L is the distance propagated through the atmosphere, and C_n^2 is a measure of the spatial structure in the index of refraction fluctuations. The low and high frequency response of the corresponding phase-difference power spectrum is

$$\begin{aligned} W_{\delta s}(f)_{low} &= 2.59 \rho^2 C_n^2 k^2 L v_{\perp}^{-1/3} f^{-2/3} \\ W_{\delta s}(f)_{high} &= 6.56 \cdot 10^{-2} C_n^2 k^2 L v_{\perp}^{5/3} f^{-8/3} \end{aligned}$$

The power law dependence at low frequencies is -2/3 and at high frequencies is -8/3. The complete spectrum $W_{\delta s}(f)$ is shown in Fig. 8.2.

Fried's coherence length r_0

The high frequency response is often expressed in terms of r_0 , Fried's coherence length. Fried (1965) defines r_0 to describe the phase structure function, such that

$$D_s(\rho) = 6.88 \left(\frac{\rho}{r_0} \right)^{5/3}. \quad (8.7)$$

From Tatarski's work we have (1961, Chapt. 8, Eq. 8.22)

$$D_s(\rho) = 2.91 k^2 \rho^{5/3} \int_0^L C_n^2(\mathbf{r}) dx, \quad \rho \geq \sqrt{\lambda L},$$

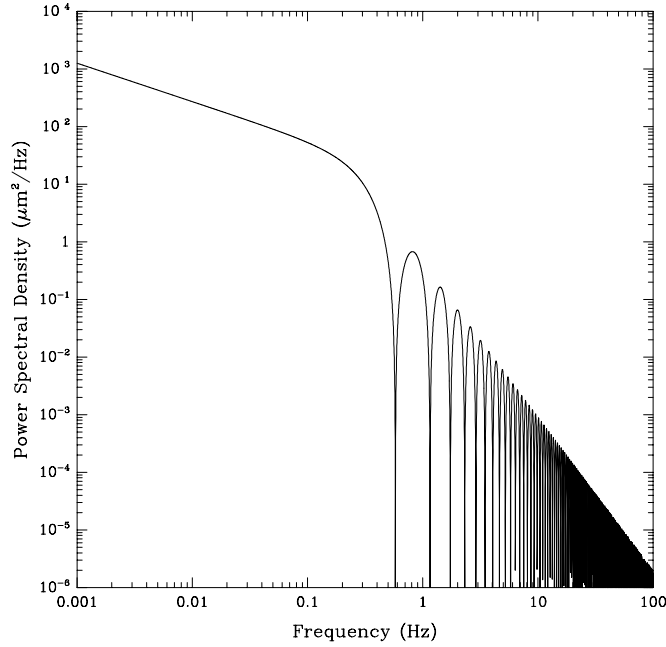


Figure 8.2: The power spectrum of phase-difference fluctuations. This example is for a baseline B of 10 m, $r_0 = 10$ cm, $T_0 = 14$ ms, and $\lambda_0 = 0.5 \mu\text{m}$. At high frequencies the spectrum is dominated by the $-8/3$ power law. At low frequencies the spectrum flattens off to a slope of $-2/3$. It has been assumed here that there is an infinite outer scale length to the turbulence.

and so r_0 can be expressed as

$$r_0^{-5/3} = 0.42 k^2 \int_0^L C_n^2(h) dh.$$

Now, for propagation through a homogeneous medium we can write

$$C_n^2(h) = C_n^2, \quad \text{and} \quad \int_0^L C_n^2(h) dh = C_n^2 L,$$

and therefore

$$W_{\delta s}(f) = 0.157 k^2 \left(\frac{v_{\perp}}{r_0} \right)^{5/3} f^{-8/3}. \quad (8.8)$$

Buscher et al. (1992) write Eq. 8.8 in a slightly different form. They combine the parameters r_0 and v_{\perp} into a single coherence time T_0 ,

$$T_0 = 0.81 \left(\frac{r_0}{v_{\perp}} \right),$$

which represents the observing time over which the rms phase change is 1 radian (Tango and Twiss, 1980, Eq. 4.9). They then use a ‘two sided transform’ to estimate the phase and express the power spectrum in terms of μm of fringe motion at a

reference wavelength λ_0 (also expressed in μm).

$$W_{\delta x}(f) = 1.38 \cdot 10^{-3} \lambda_0^2 T_0^{-5/3} f^{-8/3} \quad (8.9)$$

which may be expressed in units of $\mu\text{m}^2/\text{Hz}$.

Rms white-light fringe motions

It can be seen from the definition of the phase structure function, Eq. 8.3, that it describes rms phase variations as measured at two separated points. It may therefore be used to characterize the motions of the white-light fringe within an interferometer of baseline B . Using Eq. 8.7 and expressing it in terms of wavelengths, we have

$$\sigma_x^2(B) = 6.88 \left(\frac{\lambda}{2\pi} \right)^2 \left(\frac{B}{r_0} \right)^{5/3}.$$

By this means, Tango and Twiss (1980) predict that for $r_0 = 10\text{cm}$, and $\lambda = 500\text{ nm}$ there should be expected pathlength errors of near $B \cdot 10^{-6}$. For a Michelson interferometer operating in the visible this means that, if B is expressed in meters, the largest fringe motions should be about $3B \mu\text{m}$. At a baseline of 600 m the pathlength excursions would be greater than 1 mm.

Characteristics of the spectrum

An example power spectrum was shown in Fig. 8.2, and was calculated using Eq. 8.4. The sine term in this expression causes the nulls at higher frequencies. They are a result of the ‘frozen turbulence’ approximation: if the wind moves the turbulence by one period (on a spatial scale) then the resultant phase change is zero. The behavior at low and high frequencies is also well illustrated in this graph.

At high frequencies the spectrum is independent of the baseline. However, at low frequencies it is not, and this is an area of much speculation in stellar interferometry. Eq. 8.8 is based on a Kolmogorov model for the index of refraction fluctuations, $\Phi_n(\kappa)$. Other models exist that typically include a finite *outer scale length* for the atmospheric turbulence; among these are the Von Karman and the Greenwood-Tarazano spectra (Colavita et. al., 1987). The outer scale refers to the largest scale size at which energy is injected into the turbulent medium. It is important because it predicts the largest path fluctuations that will be present at low frequencies. Much of the analysis of the measured spectra is concerned with fitting curves to extract this parameter.

8.3 Previous observations of white-light fringe motion

There are only two interferometers that have observed power spectra of white-light fringe motions at long baselines: the Mark III interferometer and the Infrared Spatial Interferometer, both located on Mount Wilson in the United States. Although measurements have been made elsewhere, including those by Nightingale and Buscher (1992), these have been at comparatively short separations (less than 5 m).

The Mark III Interferometer

With few exceptions, the power spectra observed at the Mark III have been said to agree well with Kolmogorov theory (Colavita et al., 1987; Buscher et al., 1992). The agreement in the high frequency domain has been especially good, however at low frequencies they have obtained conflicting results. A recorded fringe motion is shown by Colavita et al. (1987) for a baseline of 12 m observing α Aql. It showed peak-to-peak changes of $16 \mu\text{m}$ over 72 seconds, and short-term variations on a scale of $5 \mu\text{m/s}$. Other power spectra, presented in the same paper, were used to suggest an outer scale length of 2 km. More recent observations, using baselines up to 31.5 m, have indicated an outer scale size comparable with the baselines of the observations (Buscher et al., 1992). This is two orders of magnitude smaller than their previous estimate. It can be seen from their graphs that the high frequency domain extends even to about 0.5 Hz, or time-scales of several seconds, and the components at low frequencies rise to about $1000 \mu\text{m}^2/\text{Hz}$ at 0.001 Hz, suggesting variations on a scale of $60 \mu\text{m}$ peak-to-peak over 15 minutes. The coherence times T_0 for observations with the Mark III have been reported to be in range of 14–26 ms at a wavelength of $\lambda_0 = 500 \text{ nm}$ (Buscher et al., 1992).

The Infrared Spatial Interferometer

The paper by Bester et al. (1992) presents measurements at the same site using an infrared heterodyne interferometer with baselines of 4 and 13 m. Unlike the Mark III observations, their data show departures from Kolmogorov theory. They state that there are undoubtedly circumstances where the theory is valid; however, typical conditions are often different from what the theory would predict, suggesting a scaling somewhere between the Kolmogorov model and a random walk. They quote coherence times, T_0 , of 10–30 ms as being typical at visible wavelengths and speculate that under excellent seeing conditions the data may support an outer scale length of 5–20 m. They are skeptical about the theory's utility and conclude that at longer baselines the pathlength fluctuations will probably not increase as quickly as has been generally believed.

8.4 Active and Passive Tracking of Fringe Motions

In the following section the performance of two different modes of fringe tracking will be examined: active and passive tracking. At high light levels it will be possible to actively follow the delay changes and compensate for them in real-time. If insufficient light were available for active tracking then passive observations could still locate the mean value of the path difference and update the delay position perhaps after several tens-of-seconds or minutes.

8.4.1 Active fringe tracking

The limitations of active tracking can be estimated if one assumes that the changes in delay occur at a constant rate over short timescales. Such changes would also be observed if the baseline orientation were not accurately known and systematic errors existed in path compensation. The delay would appear, sweep through the range of detectable values, and then vanish.

Under what conditions could active tracking be used? If the delay is moving at a rate of v $\mu\text{m/s}$, then after a time t it will have changed by vt μm . The peak in the power spectrum can be integrated provided the delay changes remain unresolved. As was discussed in §8.1.2 this will be true so long as $vt < 1/\Delta\kappa$. Therefore the total integration time should be set to $T = 1/(v\Delta\kappa)$, and sub-divided into m frames each of duration τ , where the total change is a fraction of a wavelength, α , or $v\tau = \alpha/\bar{\kappa}$. The quantity α then determines the loss factor η which reduces the fringe visibility during coherent integration, as discussed in §8.1.

$$|\gamma| \rightarrow \eta|\gamma|, \quad \eta = \frac{\sin(\pi\alpha)}{\pi\alpha}.$$

It follows that if the parameters are chosen according to these rules, as the speed v approached 0 the coherent integration time would approach ∞ . Furthermore, the maximum number of frames would always be independent of v ,

$$\frac{T}{\tau} = m = \frac{\bar{\kappa}}{\alpha\Delta\kappa}. \quad (8.10)$$

Recall that the inequality of Eq. 7.20 must be satisfied to insure proper tracking. That expression is restated here:

$$N_t > \frac{8\beta^2}{m|\gamma|^2} \left[1 + \frac{\sqrt{m}}{\beta} \right]. \quad (8.11)$$

where N_t is the total number of photons recorded in a frame, m is the total number of frames integrated, $|\gamma|$ is the fringe visibility and β is a threshold, $\beta \simeq 3$.

Let the total bandwidth extend from 440 nm to 550 nm. We have therefore that $\Delta\kappa = 0.45 \mu\text{m}^{-1}$, and $\bar{\kappa} = 2.05 \mu\text{m}^{-1}$. Now if we set $\alpha = 0.25$ then $\eta = 0.90$

and the number of frames m will be 18. We can now put these values into Eq. 8.11 using $\beta = 3.0$, and the minimum number of photons-per-frame is therefore 9.7 while $|\gamma| = 1$, independent of the rate of pathlength change v .

This arises for the following reasons: the longest integration time is restricted by the changes in delay, governed by vT ; the frame time is determined by the allowable coherence loss η and the rate v . These two restrictions limit the total number of frames that can be integrated, and for a restricted integration time there is a corresponding minimum number of photons-per-frame required to observe the delay. The number of frames is independent of v , and therefore so are the required photons-per-frame.

For delay acquisition it may be possible to improve the sensitivity of the system by deliberately reducing the resolution in the power spectrum. This could be done by averaging neighbouring samples in the spatial frequency domain: the delay would then have to travel further in order for it to be resolved in the spectrum and could therefore be integrated for longer.

8.4.2 Passive observations of delay

In the previous case it was assumed that the delay never passed through the same spatial frequency twice while being integrated, and that the purpose was to follow it as it moved. If insufficient light is available for that task, then it may still be possible to determine the mean value of the delay if the excursions are not large.

The effects of atmospheric path fluctuations were examined by Colavita and Shao (1988). They modelled the delay as being Gaussian distributed with an arbitrary mean value and a variance given by

$$\sigma_x^2 = 6.88 \left(\frac{\lambda}{2\pi} \right)^2 \left(\frac{B}{r_0} \right)^{5/3},$$

where B is the length of the baseline. This expression is based on Kolmogorov turbulence theory and was discussed in §8.2.1. They also assumed that the path fluctuations were much larger than the sampling interval in spatial frequency, $\sigma_x \Delta\kappa \gg 1$, so that sampling effects, discussed in Chapt. 6, could be ignored. After a long integration, the peak value at each spatial frequency would depend on the amount of time the delay had rested there. Changes in delay would integrate to describe a probability density function, and the central peak would have a height proportional to the area of the curve immediately below it. They then used this to estimate the signal-to-noise ratio.

Instead, let us introduce the loss factor u . Using a change of variables this can be expressed as an integration across the Normal distribution.

$$u = \frac{1}{\sqrt{2\pi}} \int_{-\epsilon}^{\epsilon} \exp\left(\frac{-y^2}{2}\right) dy, \quad \text{where} \quad \epsilon = \frac{1}{2\sigma_x \Delta\kappa}. \quad (8.12)$$

Baseline (m)	σ_x (μm)	Response u (%)
5	4.9	16.9
10	8.7	9.6
20	15.5	5.4
40	27.7	3.0
80	49.3	1.6
160	87.9	1.0
320	156.6	0.5
640	279.0	0.3

Table 8.1: The sensitivity in delay tracking for weak sources. The peak response versus baseline is tabled, assuming long integrations are required to find the mean delay. This is for $\lambda = .450 \mu\text{m}$, $r_0 = 0.10 \text{ m}$, and $\Delta\kappa = 0.477 \mu\text{m}^{-1}$. An infinite outer scale is assumed.

The peak response depends on the width of the Gaussian: the wider the Gaussian, the lower the peak. Representative values for the peak response as a function of baseline are indicated in Table 8.1. This factor can now be included in the probability of tracking loss. The power spectrum will then peak with a new value of

$$|\Lambda|^2 \rightarrow u |\Lambda|^2$$

So, taking this into account in Eq. 8.11, we have

$$N_t < \frac{1}{u} \frac{8\beta^2}{m\eta^2|\gamma|^2} \left[1 + \frac{\sqrt{m}}{\beta} \right].$$

This will substantially increase the number of photons-per-frame required to observe the group delay. However, it is possible to improve the sensitivity by reducing the resolution, $\Delta\kappa$, and tailoring it to the expected fluctuations.

8.5 Conclusion

The temporal phase-difference power spectrum for Kolmogorov turbulence was described to predict the motions of the white-light fringe that would be observed with a Michelson stellar interferometer. It was shown that for high frequencies the power spectra should have the same characteristics at all baselines. This implies that the bandwidth requirements for a fringe tracking servo will not be much different for a 100 m baseline than for a 10 m baseline. The coherence times will be limited to 10–30 ms, corresponding to rms phase changes of 1 radian. This implies that at $\lambda = 500 \text{ nm}$, changes in pathlength of 24 nm/ms may occur (3σ change).

The requirements at longer baselines will depend on the low-frequency behavior of the atmosphere. Based on Kolmogorov turbulence theory, the pathlength

fluctuations over a 500 m baseline may be as large as 1 mm peak-to-peak. However, measurements are yet to be made with baselines longer than 31.5 meters, and the limitations of the theory are not well known. Judging by current observations it is probable that variations at least as large as 10–100 μm will be present, with changes as fast as 5 $\mu\text{m/s}$ occurring over timescales of seconds.

The theoretical performance of group delay tracking has been outlined. The limitations of coherent and incoherent integration of delay have been reviewed, and a figure of merit has been derived for the threshold of delay detection for moving fringes under low light level conditions. The limitations have been spoken of in terms of the ‘number of photons-per-frame,’ where a frame refers to the photons collected across all spectral channels during the coherence time of the atmosphere. The sensitivity of delay tracking will therefore also depend on the optical efficiency of the interferometer and the detective quantum efficiency of the array detector.

8.5.1 Active vs. passive interferometry

The number of photons per coherent integration time, T_0 , limits the ability to actively track the delay location. This was discussed in §8.4.1. For the example that was considered, with a bandwidth of near 100 nm, it was found that under the best of conditions, with $|\gamma| = 1.0$, more than 9 photons-per-frame are required to actively track fringes. At lower light levels the observations are restricted to determining the mean value of the delay—which is only possible if the instrument is sufficiently stable. Shao and Colavita (1992) have designated the two modes of operation as *active* and *passive* interferometry. In the later case the interferometer must be able to track sidereal motions to a greater precision than the atmospheric fringe wander. The fringes in the channeled spectra will then oscillate about a mean spatial frequency, which can be determined over long integration times. The limitations for the passive observation of delay were outlined in §8.4.2.

8.5.2 Group delay tracking and passive interferometry

Under low light conditions the fringe amplitude detected by group delay tracking is sensitive to the instrument stability. When there is enough light the delay can be detected before it has moved from one spatial frequency bin to another. However, if long integrations are required then the signal becomes spread across many frequency bins. The peak is reduced by the height of the probability distribution of delay movements. For a total bandwidth of ~ 100 nm, centered on $\lambda = 450$ nm, a motion of ± 20 μm would smear the peak across ± 10 spatial frequency bins, but move over only $\pm 1/25^{\text{th}}$ the width of the coherence envelope. By contrast, it should be noted that the method of envelope-tracking, described by Davis and Tango (1986), is largely

insensitive to pathlength fluctuations, and is capable of observations in the presence of even larger instrumental delay excursions.

As was described in Chapt. 2, envelope-tracking is a passive method of observation. One photomultiplier is used in each of the combined beams for single wavelength observations. Narrow bandwidths are required so that the delay envelope is broad; with a bandwidth of 0.4 nm the coherence length would be near 500 μm . Observations of visibility-squared require only that the pathlength be stable to within a fraction of the width of the delay curve. Measurements are taken across the envelope at different steps in delay. Fluctuations in pathlength cause the measurement point to be ill defined, so that what is recorded is an *average* visibility over that region of the delay curve. However, unlike group delay measurements, the visibility is not substantially reduced by path changes.

If the delay is known then envelope measurements can be made at a single fixed position with respect to the peak of the delay curve, and it is no longer necessary to step through the coherence envelope to determine the peak response. The relationship between the measured and the peak visibility can be inferred from the path-difference. The calculation of visibility would remain the same as described in §2.4, but would be carried out at larger bandwidths and therefore have an increased sensitivity.

Part IV

Experiments in Delay Tracking

Chapter 9

Data Processing

Part IV of this thesis is devoted to experiments in group delay tracking using the PAPA camera. The hardware and software is described here in Chapt. 9 and the experimental results follow in Chaps. 10 and 11.

The discussion of the signal processing has been divided into several topics: data acquisition, calculation of the periodogram, integration of power spectra, and extraction of the signal peak. The timing of the electronics will not be discussed, but the data flow will be described through each component in the system. The circuit diagrams for the hardware and more detailed information about the processors have been documented elsewhere (Lawson, 1993).

9.1 Data Acquisition

The Data Format

Each photon address is represented by a pair of 8-bit numbers, x and y , but only the x coordinate is used in processing. It is assumed that all photon events issued by the camera belong to the channeled spectrum, and to insure this is true a slit is placed at the image intensifier. It is aligned along the x axis so that only photons from the spectrum will be seen by the camera's optics.

The same process could also have been done electronically, by latching the y data and rejecting events that occurred away from the spectrum.

The Data Acquisition

As shown in Fig. 9.1, the 8-bit data are integrated in static RAM (SRAM). There are two banks of memory, each with 512 registers. The lower half of the memory is

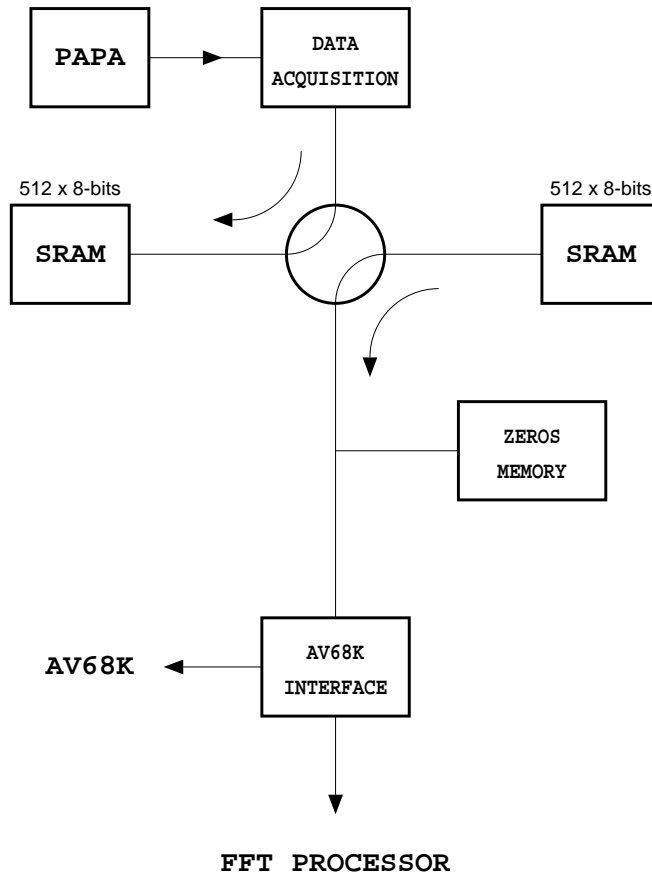


Figure 9.1: The data flow in the Data Acquisition module. The data from the PAPA camera is integrated in one of two buffers of static RAM, while the other buffer of just-collected data is read out to the FFT processor. The memory of each buffer is reset to zero after its data has been read out, and the two buffers swap after each integration period. The buffers are single mapped to the address space of the AV68K 68000 computer.

used for data integration, while the upper half is used for a look-up table. When a photon arrives its address is used as an index to the table, and an appropriate location in memory is incremented. This continues until the end of the integration period, which will be between about 1.5 and 10 ms duration depending on the atmospheric seeing conditions. A ping-pong buffer arrangement ensures that all photons events are integrated: while one buffer is filling up with data the other is being processed, and once the data had been removed the memory is erased. At the end of the cycle the two buffers swap and the cycle repeats. Each bank of memory is mapped into the control computer, and the look-up tables can be easily changed in software.

Although the hardware could run with a period of 0.2 ms, the shortest integration time had to be greater than 1.2 ms, since that was the time it took for the

interrupt routine to process data.

9.2 The Calculation of the Periodogram

In each cycle the periodogram of the integrated data is calculated. A 256-point FFT is performed with integer arithmetic using an Austek 41102 Frequency Domain Processor. The transform of the data is produced from the processor as a series of complex numbers (x_m, y_m) . The modulus of each number is then calculated by a Plessey ‘Pythagoras Processor’ $r_m = (x_m^2 + y_m^2)^{1/2}$, and the result is written into an array in memory.

Austek A41102 Frequency Domain Processor

The A41102 is a CMOS pipeline processor used for computing FFTs in real-time with up to 256 points, and may be cascaded to perform much longer transforms. It operates in 16, 20, or 24 bit precision, and will perform a 256-point transform in 102.4 μ s when used in 16-bit mode. The architecture of the chip will not be described here. More detail about the A41102 processor can be found from its *User Manual* (Austek, 1988a) and *Product Specification* (Austek, 1988b). The terminology and algorithm of the FFT are described by Bergland (1969) and Cochran et al. (1967).

The photon data arrives as a series of 8-bits samples, $h(n)$, which are read into the A41102 as if they had been multiplied by 2^{12} . The data ports on the chip are 24-bits wide, and are illustrated at the top of Fig. 9.2. The calculations are performed within the chip using 20-bit precision, with the top 16-bits used in the next stage of processing. The four least-significant bits are not used. The transform can be written as

$$\text{Re}[H(m)] = \sum_{n=0}^N \left\{ 2^{12} \cdot h(n) \right\} \cos\left(\frac{2\pi nm}{N}\right), \quad 20 \text{ bits} \quad (9.1)$$

$$\text{Im}[H(m)] = \sum_{n=0}^N \left\{ 2^{12} \cdot h(n) \right\} \sin\left(\frac{2\pi nm}{N}\right), \quad 20 \text{ bits} \quad (9.2)$$

where $N = 256$. The real and imaginary parts are output as integer 16-bit numbers

$$x_m = 2^{-4} \text{Re}[H(m)], \quad 16 \text{ bits}, \quad (9.3)$$

$$y_m = 2^{-4} \text{Im}[H(m)], \quad 16 \text{ bits}. \quad (9.4)$$

The A41102 has three data ports which can be arbitrarily configured as input or output. Port A is used for input, Port C for output, and Port B is unused. The chip is software programmed through its Supervisory Port to perform 256-point forward transforms, reading in the data samples in bit-reversed order, with output in natural

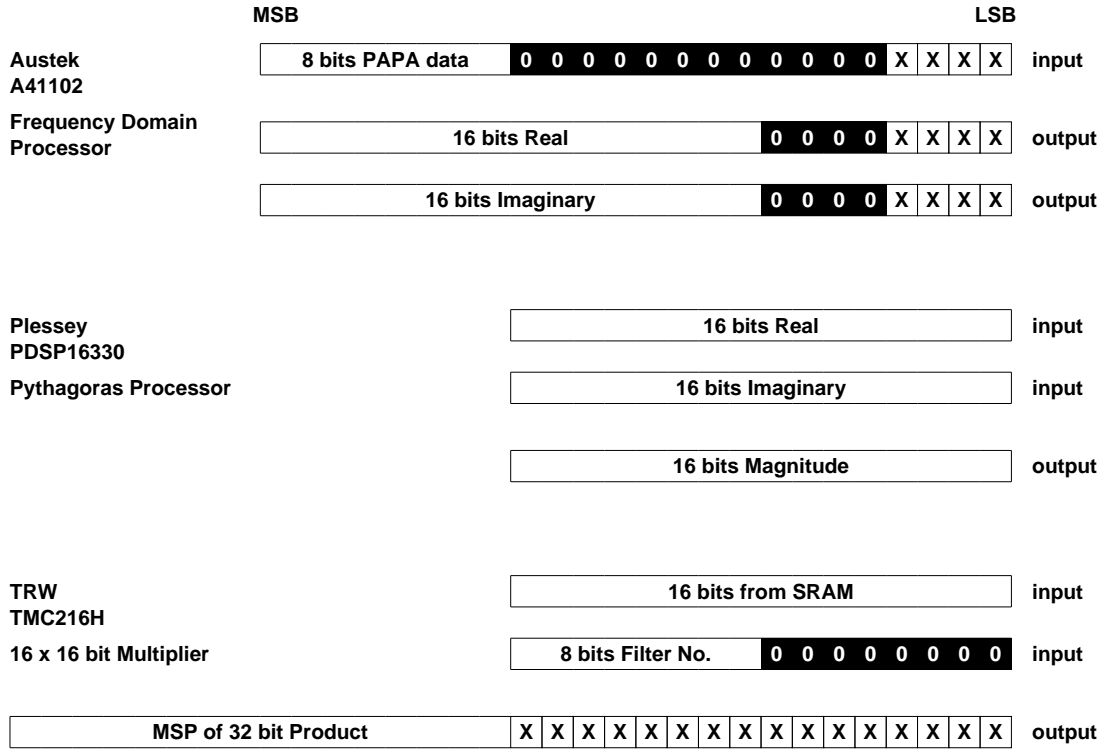


Figure 9.2: The data flow through the processors. Integer arithmetic is used throughout. The PAPA data is effectively multiplied by 256 on entry to the Austek Fourier Transform Chip. The amplitude of the transform is then calculated in 16-bit precision by the Pythagoras Processor. The last stage in the process is an iterative integration, which may be performed using the TRW multiplier: it reads in a 16-bit number and outputs a 16-bit number which may then be added to incoming data from the Pythagoras processor. The addition, which is not shown here, is also performed with 16-bit precision.

order. Scalers are available at each butterfly in the FFT, allowing the data to be divided by factors of 2 to avoid integer overflows. The scalers may be enabled by setting the appropriate bits in the Supervisory Port, and can be used to improve the accuracy of the calculation. The effect of using integer arithmetic is covered in detail by Welch (1969).

Plessey PDSP16330, Pythagoras Processor

The PDSP16330 is a digital CMOS IC that performs cartesian to polar coordinate conversions in 16-bit precision at high speeds. 16-bit x_m and y_m data can be clocked into the chip every 100 ns, and 16-bit magnitude r_m and phase ϕ_m will appear at the output 24 clock cycles later.

$$r_m = (x_m^2 + y_m^2)^{1/2}, \quad 16 \text{ bits.} \quad (9.5)$$

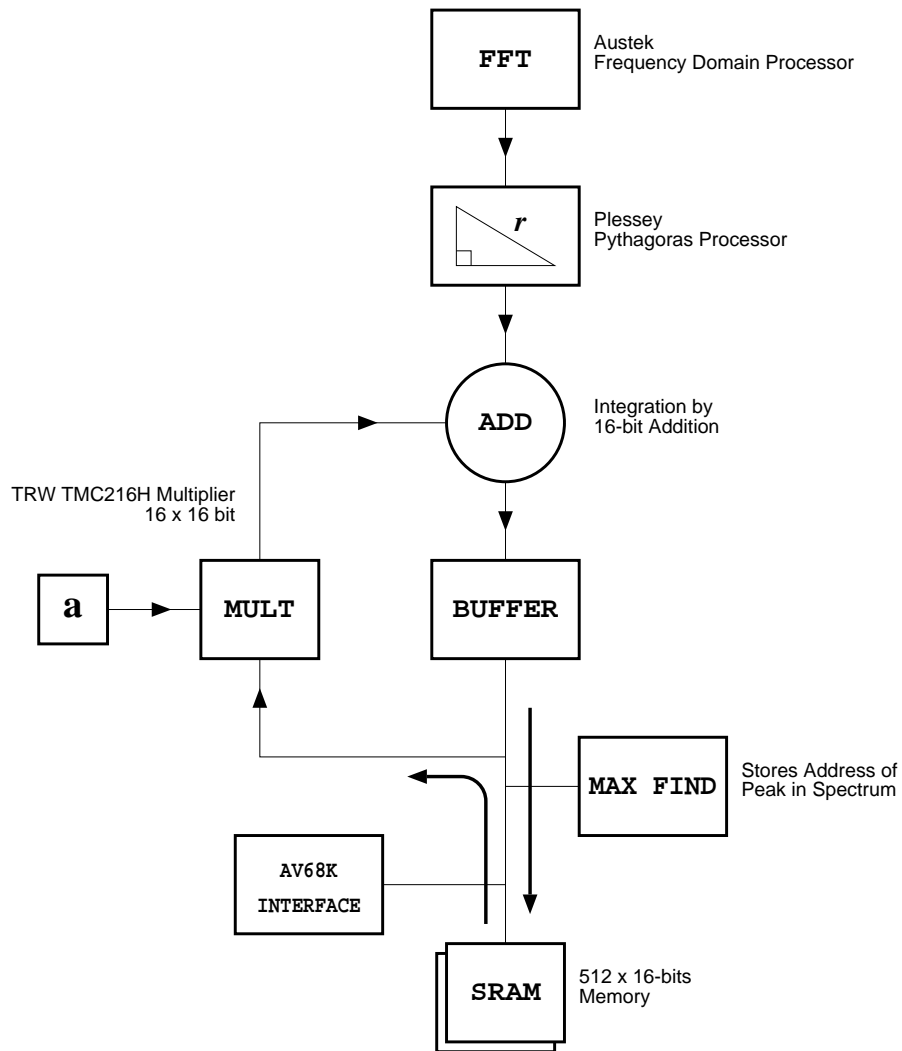


Figure 9.3: The data flow in the FFT processor module: The data is read into the FFT chip and then passed through the Pythagoras Processor. This processor does a cartesian to polar coordinate conversion, and the modulus of the transform is stored in memory (the phase information is not kept). This spectrum is multiplied by a constant less than 1 and then added to the next incoming spectrum, making it possible to integrate weak signals. The address of the peak in the spectrum is determined as the spectrum is written into memory.

The phase information is ignored. The operation of this chip is described by Plessey (1988). It accepts two's complement x_m and y_m input, and is therefore compatible with the output of the A41102.

9.3 Integration of Power Spectra

Provision was included in the electronics to recursively integrate the power spectra in the manner described by Bozic (1979, §6.3). The method of integration is illustrated by the flow diagram of Fig. 9.3. After the data leaves the Pythagoras Processor it is summed with a fraction of the previous spectrum and is then written into memory. The input-output relationship is described by the difference equation

$$y(n\tau) = a y(n\tau - \tau) + x(n\tau), \quad a < 1. \quad (9.6)$$

In this formulation $y(n\tau)$ is one component in a new estimate of the power spectrum, $y(n\tau - \tau)$ was the previous estimate stored in memory, and $x(n\tau)$ is new data from the Pythagoras Processor. The recursion retains information from previous samples, over a time constant determined by the choice of a . Its frequency response function is given by Rader and Gold (1967),

$$|H(f)| = \left[1 + a^2 - 2a \cos 2\pi f\tau \right]^{-1/2}, \quad (9.7)$$

and is plotted in Fig. 9.4. This filter is said to have an Infinite duration Impulse Response (IIR). The coefficient a determines what fraction of the previous samples contribute to the current estimate. When $a = 0$ all previous samples are forgotten, but when $a = 1$ then all previous samples are retained. This is also illustrated by the frequency response: as the value of a is increased the response to higher frequencies is reduced and the filter integrates for longer. If the new data is roughly constant with each new measurement, $x(n\tau) = x$, then after a long time the recursion will yield approximately

$$y(n\tau) \simeq \sum_{n=0}^{\infty} a^n x = \frac{x}{1-a}, \quad 0 \leq a \leq 1. \quad (9.8)$$

The mean level of the power spectrum can be derived from the filter coefficient and the current estimate:

$$\bar{x} = (1-a) y(n\tau). \quad (9.9)$$

TRW TMC216H, 16 x 16 bit Multiplier

The TMC216H is another high speed CMOS chip. It performs a 16 x 16 bit multiplication in 145 ns, and accepts a variety of input data formats. It is used as the

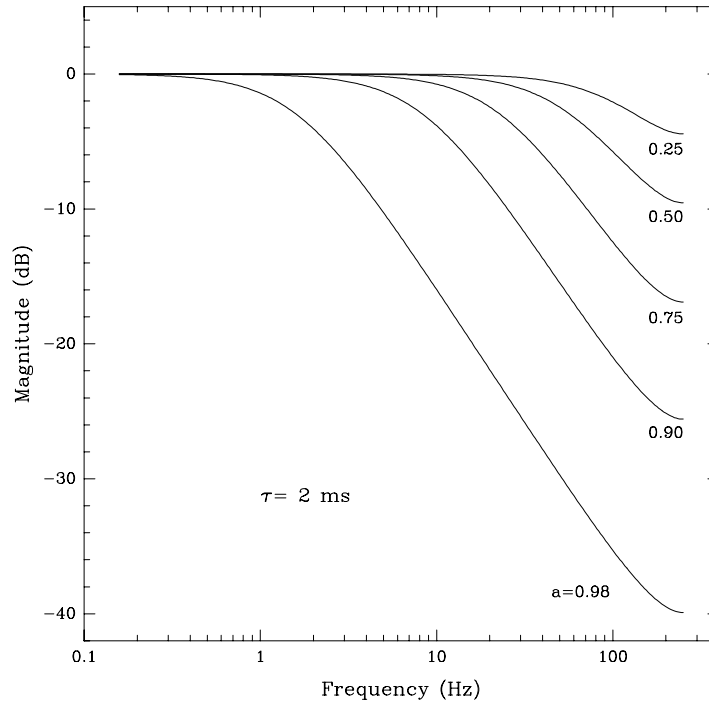


Figure 9.4: The frequency response of the single-pole recursive filter for $\tau = 2$ ms and with various filter coefficients, a .

multiplier in the recursive filter. The 16-bit $y(n\tau - \tau)$ data is input from memory and effectively multiplied by the fraction

$$a = \frac{\mathbf{a}}{256}, \quad (9.10)$$

where \mathbf{a} is an 8-bit integer: the data is first multiplied by $\mathbf{a} \cdot 2^8$ and then divided by 2^{16} . The result is later added to incoming data from the Pythagoras Processor and written back into memory. The multiplication is illustrated in Fig. 9.2. Only the top 16-bits of the product are used, so the division by 2^{16} is performed automatically.

9.4 Locating and Filtering the Delay

9.4.1 Locating the peak in the power spectrum

The integrated spatial frequency spectrum will contain a peak at the zero-frequency that is proportional to the total number of photons recorded. It will also have low frequency components that describe the shape of the stellar spectrum, and a peak at the spatial frequency of the channeled fringes.

The location of this peak is determined during the recursive integration. As each component in the spectrum is being stored in memory, it presents itself on a data bus that is monitored by a ‘Max-find’ module, shown in Fig. 9.3. The module ignores the first low-frequency components, but latches the 10th one. Each subsequent component is compared to see if its amplitude is higher. If so, then it replaces the previous maximum, and both its amplitude and address are recorded. This continues until the whole array has been stored in memory, whereupon the peak in the spectrum will have been found. The peak is determined during every cycle of the integration.

9.4.2 Tracking the delay

The filter coefficient a is set to correspond to the bandwidth of the expected path motions. If the visibility of the fringes is good and the fringe peak stands out, then it would be straightforward to filter the raw estimates and provide feedback for tracking.

However, if the signal-to-noise ratio is low then the task becomes more difficult. When the fringe amplitude has a height comparable with the noise there will be many recorded peaks located randomly, spanning the full range of spatial frequencies. Nevertheless, a subset of the peaks would produce a trace through the data which would announce the presence of a signal. This is well illustrated by the data presented in Chapt. 10. See for instance Figs. 10.2 and 10.4. The noise events may induce large and rapid fluctuations of the peak position, far in excess of the signal motions. The delay tracking should therefore be able to do several tasks:

1. Identify the presence of a signal.
2. Selectively ignore likely noise events.
3. Use the ‘good’ data to estimate the current delay.

Let us consider each of these steps in turn.

The detectability of a signal

If there is no signal present then the peak location in the spectrum would change randomly. A plot of event location versus time would show a uniform distribution of events across all spatial frequencies. The mean *change* in delay from one sample to the next would be large, and so would the bandwidth of delay changes.

If a signal is present and distinguishable, then on average its height will be higher than the noise, and it will be detected as the global peak more frequently. The trace of peak location versus time would show clumping in the data, and the average position shift from one peak to the next would be reduced. If such a trend suggested the data was less random, then it would indicate the presence of a fringe.

A running mean of the *change* of peak location is a simple measure of the randomness of the data. If the past locations are given by $p(n - k)$, and the average change is now $d(n)$ then we can write

$$d(n) = \frac{1}{M} \sum_{k=0}^{k=M} |\Delta p(n - k)|, \quad (9.11)$$

where

$$\Delta p(n - k) = p(n - k) - p(n - k - 1). \quad (9.12)$$

This should drop to lower values if a trend exists.

The removal of noise events

Once the signal is found it can be bracketed by a spatial frequency filter: a rectangle function centered on the spatial frequency of the signal and spanning up to 10 μm of delay either side. The filter boundaries can be re-defined after each frame, and so the filter moves with the signal. Any peaks which occur outside those boundaries will be ignored, and the signal becomes the highest peak *within* the filter. A momentary drop-out would not cause a loss of tracking: the local maxima inside the filter would be used until the signal returns.

If it becomes obvious that the peak has slipped outside, or has been lost entirely, then the filter is disabled. This will stop it wandering around aimlessly, bracketing one noise peak after another.

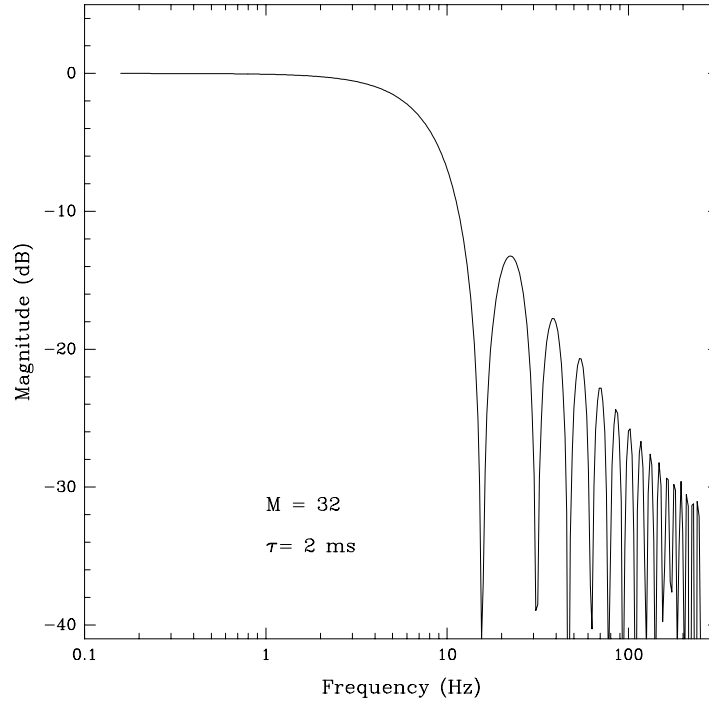


Figure 9.5: The frequency response of the moving average rectangular window. In this example $M = 32$ and $\tau = 2$ ms. The 3 dB point is near $f = 1/(2M\tau)$ and the first null is at $f = 1/(M\tau)$.

The current delay estimate

The remaining data is then low-pass filtered using a bandwidth again tailored to the expected phase-difference power spectrum. The average peak location can be determined over a period of M samples,

$$y(n\tau) = \frac{1}{M} \sum_{k=0}^{M-1} x(n\tau - k\tau). \quad (9.13)$$

This is easy to program and has the virtue of requiring only a small number of calculations in each cycle: one need only add the most recent term, $k = 0$, and subtract the oldest, $k = M$. Furthermore, if M is a power of 2 then the division can be performed by bit-shifting the sum to the right. The frequency response of this filter is shown in Fig. 9.5, and is given by Proakis and Manolakis (1992, Eq. 8.1.16),

$$|W(f)| = \frac{|\sin(\pi f \tau M)|}{|\sin(\pi f \tau)|}. \quad (9.14)$$

The value of M determines the 3 dB point of the filter, and the phase response is linear.

9.5 Passive Tracking

If 100 nm of bandwidth is used then the resolution in delay would be $2\text{ }\mu\text{m}$. From the discussion in Chapt. 8 it is reasonable to assume the delay will change by several microns per second, and therefore the incoherent integration time would be limited to several hundred milliseconds, or a few hundred frames of photon data. By inspection of Figs. 7.6 and 7.7 that implies it may be possible to actively track with as few as 10, or even 3, photons-per-frame, depending on the visibility of the fringes and the level of the background counts.

If active tracking is impossible then longer integration times are required so that the mean position of the delay can be found. Although the hardware is not capable of integrating more than about 200 frames of data, since the filter of Eq. 9.8 will overflow, the data can be transferred to the computer memory and integrated using 32-bit precision. This allows the data to be accumulated for half an hour or more.

9.6 Software and Hardware

The hardware was interfaced to a 68000 board computer, the AV68K, which is made by Interrupt Systems (1988). Clock signals are issued for every frame, and at the end of each cycle the electronics issue an interrupt request to the computer. At that time the hardware registers are read, the ‘Max-find’ location is transferred, and, if passive tracking is used, so too is the complete spatial frequency spectrum.

The control software was written in the C programming language, and the hardware registers, including the Supervisory Port of the A41102, are each mapped to specific addresses in computer memory. There are 64 different software registers which are defined for serial communication with other computers. The AV68K has two RS232 serial ports: one is used to monitor the status of the fringe tracker, and the other is used to send data to the computer of the path-compensating carriage.

It is therefore possible to send feedback signals to the carriage and close the servo loop for delay tracking. In its current implementation the carriage computer reads the data every 80 ms, and, as was discussed in §9.5, this provides sufficient response to follow the delay fluctuations. The servo equations are programmed in the carriage computer and include error signals from the laser metrology system.

9.7 Summary

In this chapter the data processing was described with emphasis on methods used for active delay tracking. The raw data are 8-bit photon addresses which are issued by the PAPA camera, and which arise in the channeled spectrum. These are integrated in a ping-pong buffer, Fast Fourier Transformed, and the resulting power spectra are then integrated recursively. The time constant for the integration is set by the value of **a** (an 8-bit number), and integer overflows are avoided by setting an appropriate number of FFT scalers in the A41102. The peak in the power spectrum is recorded by the Max-find module and is then filtered to derive the current estimate of the delay. Non-linear filtering is used to improve the performance by selectively ignoring likely noise peaks.

The processor was used for both laboratory and stellar observations, which are now presented in Chaps. 10 and 11.

Chapter 10

Performance Tests

The rebuilt PAPA camera was completed in July of 1992, following the final testing of its new analog electronics. It was used for laboratory experiments in group delay tracking through until December and was then moved to Narrabri, to be installed at SUSI. The performance of the delay tracking system is described in this chapter. The stellar observations are described in Chapt. 11.

10.1 Laboratory Measurements in Sydney

An Interferometer for channeled spectra

A Michelson interferometer was set up in the lab, as shown in Fig. 10.1. Several light sources were available for use: a HeNe laser to align the instrument and remove tilt fringes; a xenon lamp to produce the channeled spectra; and a low pressure mercury vapour lamp to calibrate the spectrometer (not shown). The light was focused on a pinhole and collimated to a 20 mm diameter beam, allowing the pinhole to be viewed as an artificial star.

In one arm of the interferometer was a piezo activated mirror with an axial movement of 40 μm , allowing pathlength changes of 80 μm . It was placed on a translation stage and mounted so that its angle of tilt could be adjusted. A programmable function generator was used to drive it and to produce sinusoidal pathlength changes. The DC level of the sinusoid could be adjusted at the piezo controller.

The output of the interferometer was directed to a prism of SF52 glass, which had been rotated to the angle of minimum deviation. The dispersed light was then focused to the input of the PAPA camera, using a 400 mm focal length achromat to image the spectral band from 400–520 nm. The output of the camera was then sent to the processor described in the previous chapter. In these examples the data were

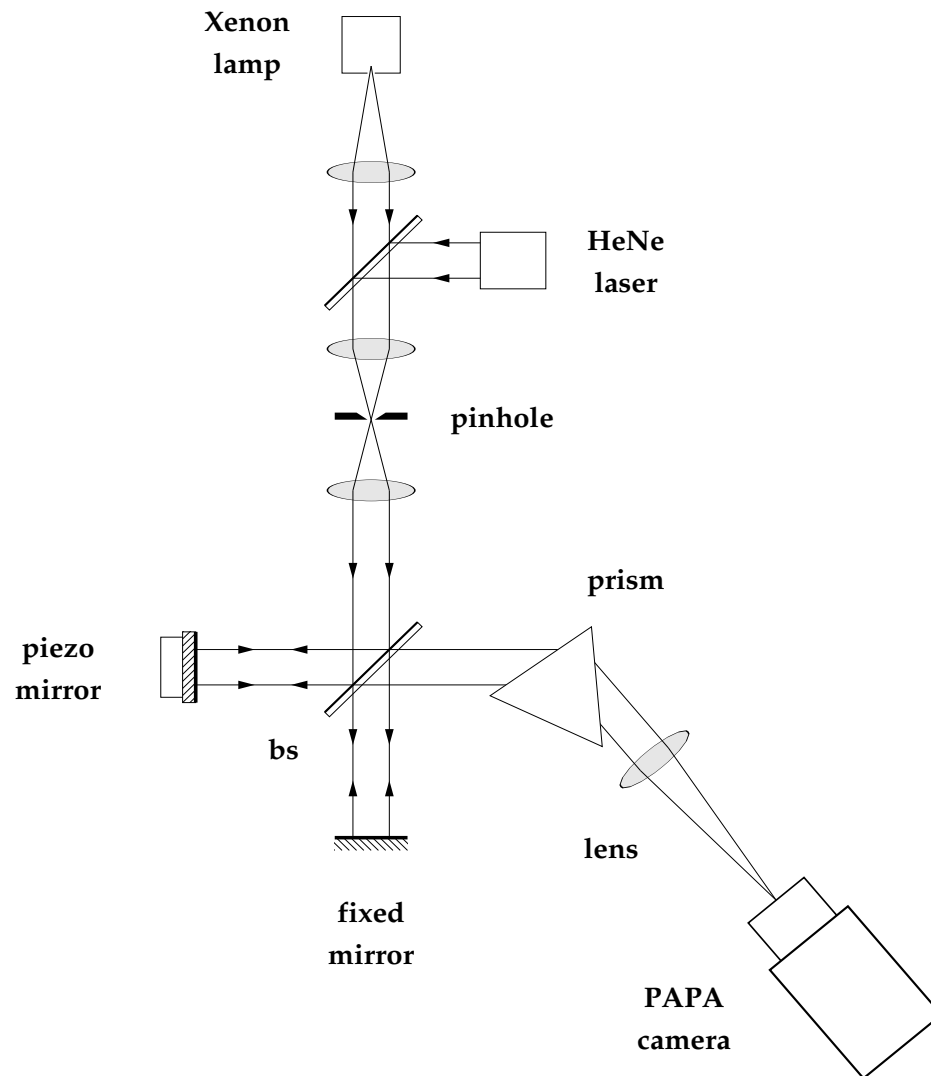


Figure 10.1: The layout of the Michelson interferometer used for laboratory tests. Not shown here are neutral density filters, polarizers, the low pressure mercury vapour lamp which could replace the laser, and the compensating plate which was included in one arm of the instrument.

averaged to 128 pixels and zero-padded prior to the FFT. A 256-point transform was used yielding 128 spatial frequency components each representing a change of about $1\text{ }\mu\text{m}$ optical path difference.

Observations of pathlength changes

It was found that for high light levels, with 100 or more photons-per-frame, the processor could observe pathlength changes with peak-to-peak amplitudes of $20\text{ }\mu\text{m}$, at frequencies up to 3–4 Hz. At higher frequencies and larger amplitudes the fringe would disappear where the fastest delay changes occurred. It would be lost for part of the period, but would re-appear at the top and bottom of the sine wave. It was still possible to find the delay with as few as 10 photons-per-frame, providing the period and amplitude of the sinusoid were reduced.

Figure 10.2 shows a trace of a 2.5 Hz sine wave with a peak-to-peak amplitude of $26\text{ }\mu\text{m}$. There are 2000 frames represented. Each dot in the figure is the recorded peak for a single estimate of the delay, as seen by the Max-find module (discussed in §9.4.1). There is scattered noise, but the location of the delay is easy to see.

At low light levels the signal-to-noise ratio was generally poor, and so the filtering methods discussed in §9.4 were implemented. This is illustrated by the data in Fig. 10.3, which was processed at the same time, but ignored the noise peaks. In each interrupt cycle the software recorded the location of the global peak, detected by the Max-find module, and noted the delay-change from one estimate to the next. If the change were to cause a jump greater than $10\text{ }\mu\text{m}$, then the most recent estimate was ignored and a local maximum was used. All of the widespread noise was thereby removed. There are however, two locations where the tracking was lost because the signal slipped out of the filter. The filter then followed the local maxima and jumped from one noise peak to another until the true signal passed within range.

Figures 10.4 and 10.5 show similar data sets. The amplitude of the delay changes is now larger, and the signal-to-noise ratio is lower. A slightly different pre-filter was used. In this case, if the tracking was lost the filter would increment the delay towards the global maximum at each interrupt cycle. It would therefore not wander aimlessly and would be faster to regain the tracking.

By these tests it was shown that the processor could follow the delay even under poor conditions, and that it would be capable of tracking atmospheric path motions. The tracking would not fail until the pathlength changes were larger than those that could be induced by the atmosphere. An amplitude of $20\text{ }\mu\text{m}$ peak-to-peak at 2.5 Hz causes a delay change of $157\text{ }\mu\text{m/s}$ at the zero crossing of the sine wave. In these tests a frame time of 1.5 ms was used, and so the most rapid change was 240 nm per frame. As was discussed in Chapt. 8, the atmosphere will cause rms phase changes on the

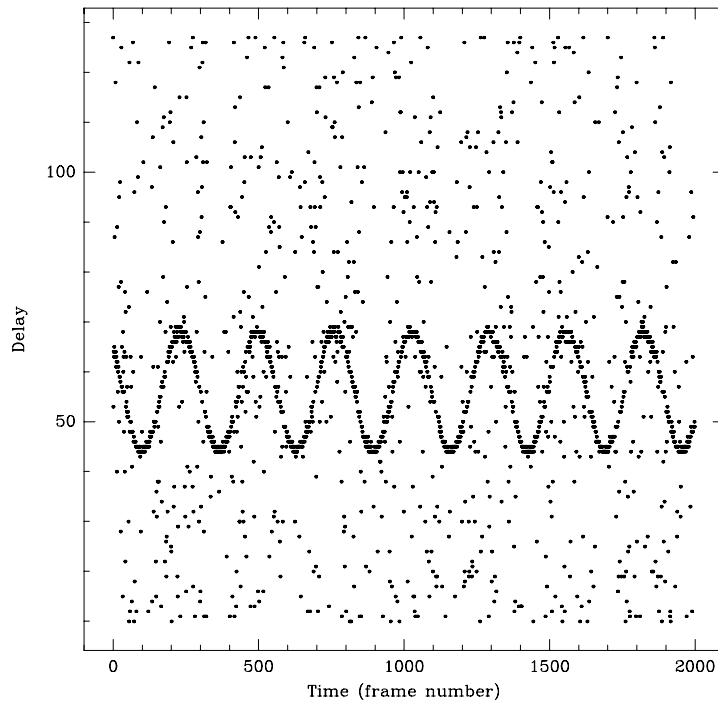


Figure 10.2: The response of the processor to a sinusoidal delay change of $26\ \mu\text{m}$ peak-to-peak and a frequency of 2.5 Hz over 3 seconds. The delay axis is $1\ \mu\text{m}$ per division. The resolution in tracking is $2\ \mu\text{m}$.

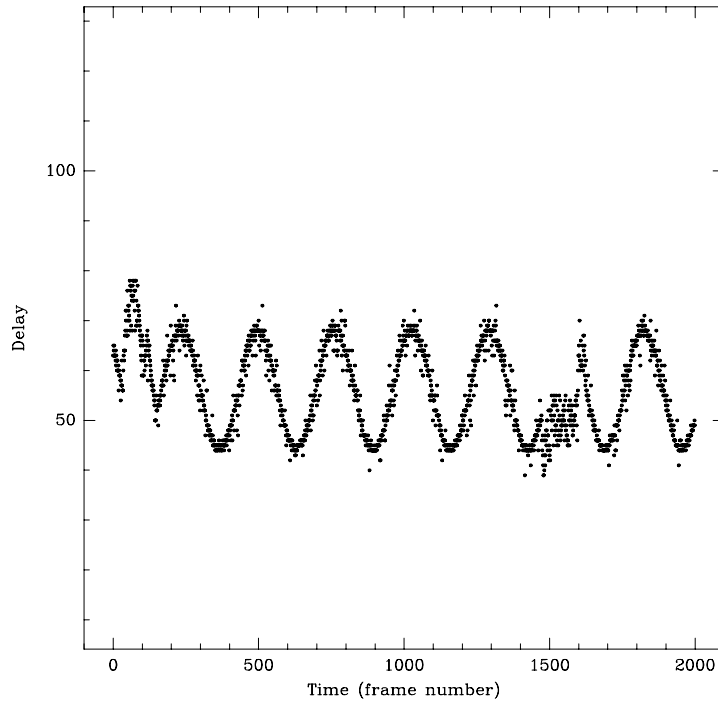


Figure 10.3: The data have now been passed through a rectangular filter that ignores events more than $5\ \mu\text{m}$ away from the last estimate.

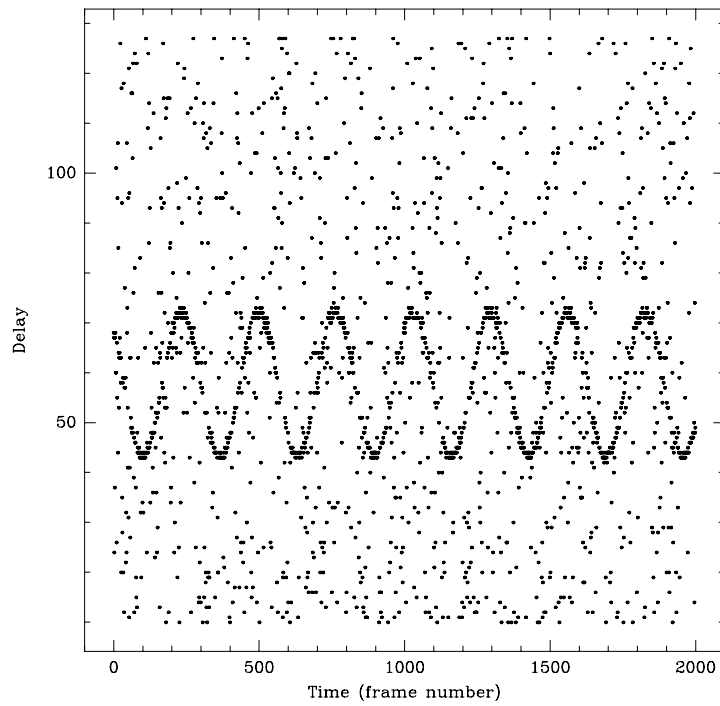


Figure 10.4: Another trace of delay with a higher amplitude.

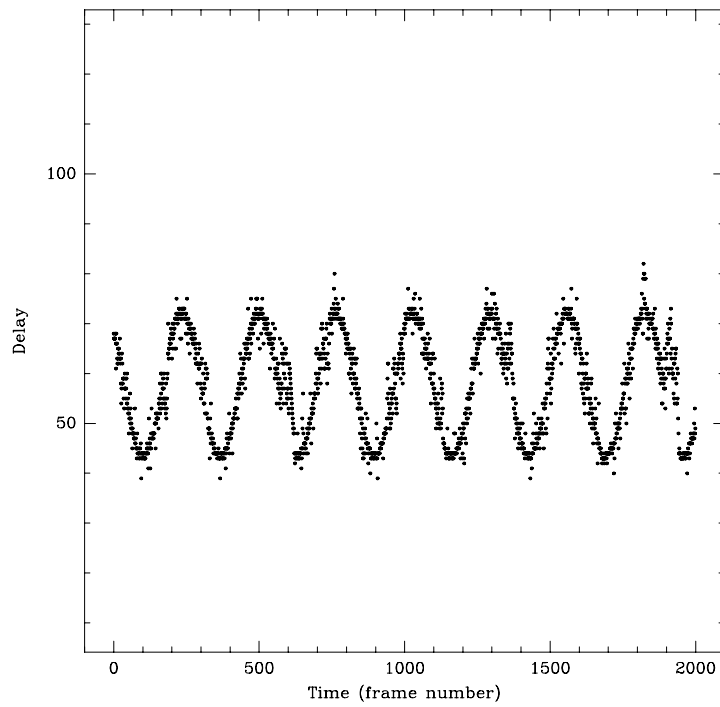


Figure 10.5: The previous trace after being passed through a rectangular filter which travels towards the current estimate if the tracking is lost.

order of 1 radian per 10–30 ms, or only 36 nm measured over the same frame time (3σ change at $\lambda = 500$ nm).

The high frequency response of the delay tracking would be limited to ~ 5 Hz, since changes of less than $2\ \mu\text{m}$ cannot be resolved by the processor, and it would take several hundred milliseconds for the atmosphere to change the path length by that amount. Much larger path changes may be induced by the atmosphere, but since they would occur at lower frequencies they would not be difficult to track. The PAPA camera and group delay hardware were therefore installed at SUSI.

10.2 Tests using SUSI in Autocollimation

In this section experiments are described to evaluate the processor's response to large delays and to measure the noise level in the DFT amplitude estimates. The following tests were performed with SUSI configured for autocollimation. In that mode SUSI acts as a standard Michelson interferometer, but with arms more than 140 m long.

There are two large optical tables within the main enclosure of the instrument that are intended for separate observations in the blue and red parts of the visible spectrum. When the camera was taken to Narrabri only the blue table was being used for visibility measurements. The combined beams could be redirected to the red table where the prism spectrometer and PAPA camera were installed. Two periscopes were used to redirect the light, which was removed after the beam combiner but before the prism that is labeled S_2 in Fig. 1.1.

Autocollimation from the Beam Reducing Telescope

SUSI is routinely aligned by autocollimation prior to an observation. An overview of the alignment procedure is given here to show how channeled spectra were produced for tests with the camera. The beam division and recombination is shown in Fig. 10.6. Light from a HeCd laser is injected at the beam combiner (bc) so that two beams are sent out through the instrument, one towards the north end of the baseline and one toward the south. As can be seen in Fig. 1.1, the beams travel the length of the optical path compensator before reaching the beam reducing telescope. An autocollimating mirror is placed there to reflect both beams back along their paths, returning them to the beam combiner.

There are three quadrant detectors on the main table: north, south, and reference. The reference detector receives starlight after the beam combiner and is used to align the north and south beams so that their wavefronts are parallel. The north and south detectors look into each beam separately, and are used as the detectors in the tilt correcting servo. Each beam is centered on its quadrant detector and then

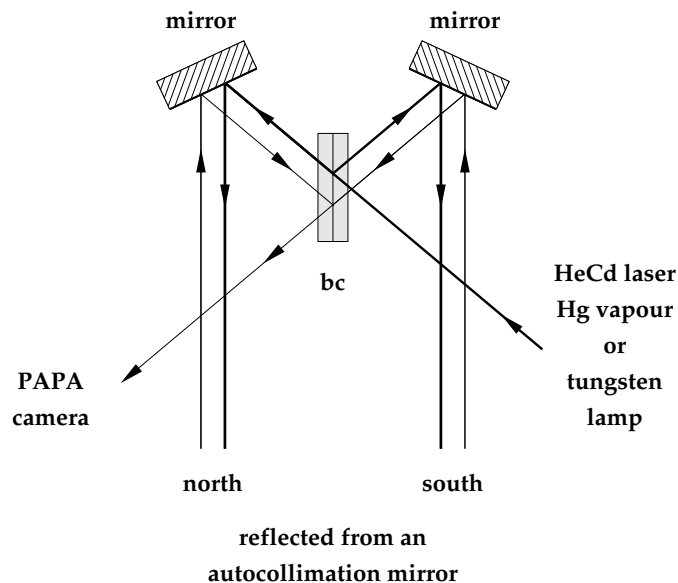


Figure 10.6: Beam combining for autocollimation with SUSI. The main beam combiner (bc) is shown. The light is injected from the right and is divided into two beams which travel the length of the optical path compensator to the Beam Reducing Telescope. They are then reflected back by an autocollimating mirror, return to the beam combiner, and exit on the left. An HeCd laser is used for alignment, and a tungsten lamp is used for the experiments with channeled spectra.

compared with the reference detector. Shutters are used during an alignment to select which beam reaches the reference detector, and adjustments are made by changing the tilt of selected mirrors. When the two beams are each centered on their respective quadrant detectors *and also on the reference detector*, then the two wavefronts will be parallel.

Further fine adjustments are required to view channeled spectra. A video monitor is used to observe tilt-fringes in the recombined laser light, and fine adjustments are made to the tilt mirrors to remove the residual tilt between beams. The light source is then changed from the HeCd laser to a mercury vapor lamp, and the fine carriage is positioned so that channeled spectra can be observed visually near the brighter spectral lines. This is performed with the aid of a microscope. It is then possible to step the carriage to vary the frequency of the channeled spectra near the white-light fringe position. As a final step the light source is replaced by a tungsten lamp which has a broad, faint, and featureless spectrum. The interferometer is now ready for the measurements.

The pathlengths can be changed by moving a piezo at the focus of the catseye on the path compensator carriage. Each pass through the carriage adds twice the piezo motion, and in autocollimation this is doubled again. The 30 μm piezo was used to provide path differences up to 120 μm .

Response at large path-differences

The design of SUSI does not allow the tilt correcting servos to be run in the autocollimation mode, and seeing within the enclosure can only be corrected while observing a star. The seeing was noticeable during the tests and resulted in variations in fringe visibility. Measurements were therefore restricted to instants when the seeing was apparently good.

The photon data were averaged to 128 pixels before the periodogram was calculated, and the bandwidth was the same as in Sydney, so each step in spatial frequency corresponded to 1 μm path difference. Figure 10.7 shows four power spectra at different delays. The DC component and the peak are obvious in each case, and the signal is easily identifiable out to 60 μm . At larger path differences the peak rapidly deteriorates until at 100 μm it is barely visible above the noise. Figure 10.8 is a composite of many similar measurements with each peak representing a different power spectrum. The path difference was stepped from -60 to +60 μm . Measurements were therefore taken either side of zero path difference, and ‘negative’ spatial frequencies would be meaningless, but the data are plotted this way to indicate the path difference was negative. Subsequent data sets were then taken to show path differences from +60 to 120 μm .

The effect of integrating the power spectra was also examined. The number of frames was increased and records of the power spectra were taken. This was done with and without the remapping of photon data. It can be seen in Fig. 10.9 that if enough frames are integrated the noise disappears and the signal can be seen at spatial frequency of 12, corresponding to a delay of 24 μm . There is also a peak at 96 which is due to alignment artifacts in the camera. This peak is removed when neighboring pixels are averaged, as in Fig. 10.10, and the signal peak now appears at a spatial frequency of 24, as expected. The calibration from spatial frequency to delay depends on the bandwidth and extent of the pixel averaging, as was discussed in §6.1.3.

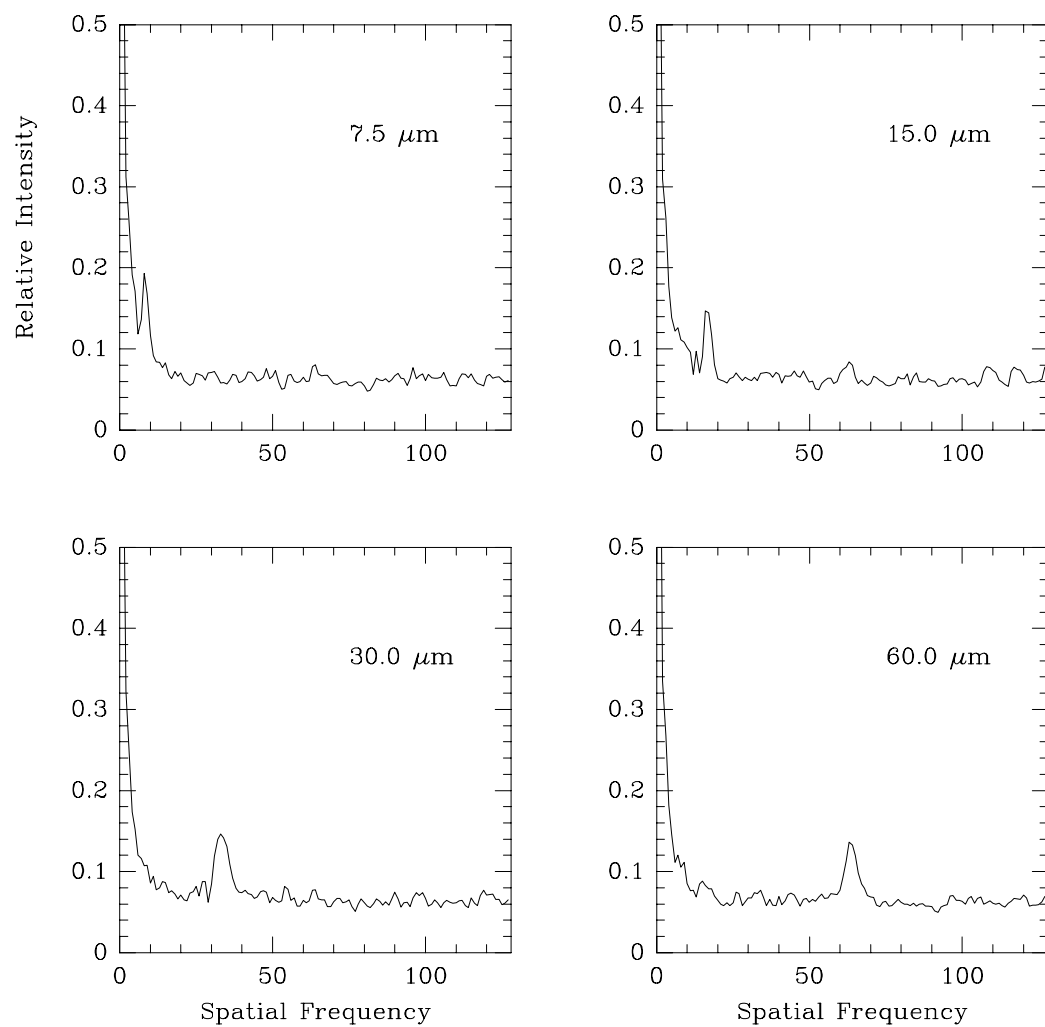


Figure 10.7: Examples of the fringe peak for increasing pathlength. The distances which are indicated in the diagrams are with respect to the center of the piezo motions and not with respect to the location of zero path difference.

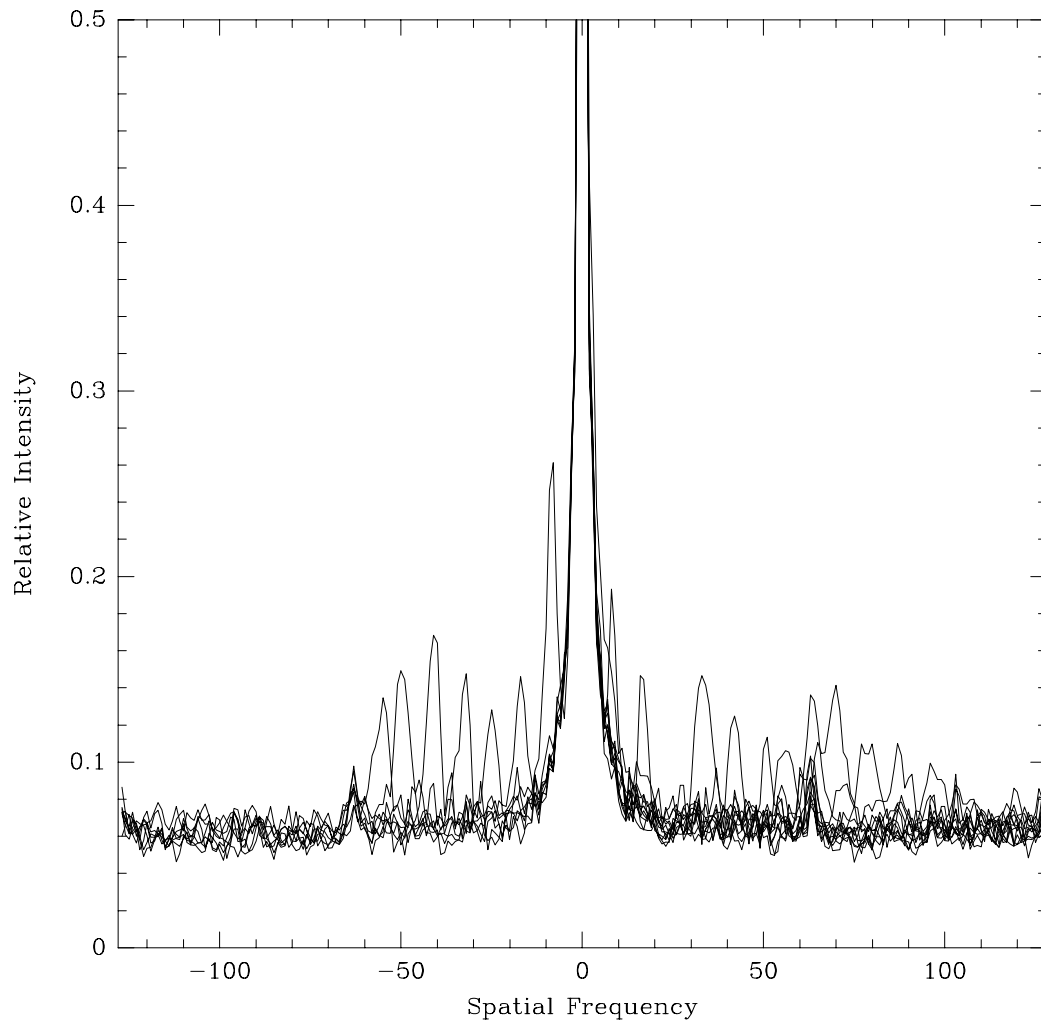


Figure 10.8: Fringe measurements made using a tungsten lamp within the stellar interferometer. The data have been averaged to 128 pixels and zero padded prior to being transformed. The height of the peaks is variable due to tilts induced by air motions within the enclosure of the path compensator.

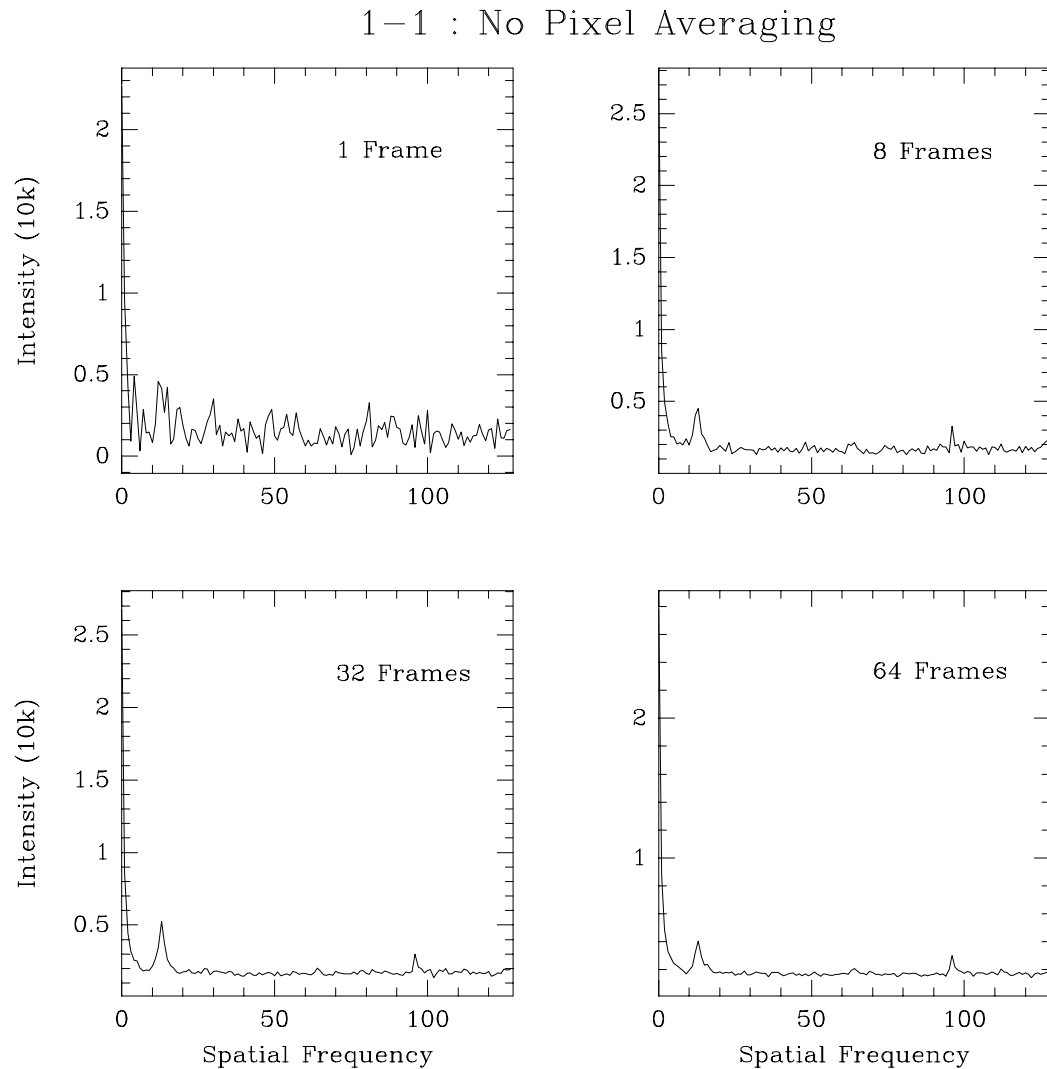


Figure 10.9: The above data sets were obtained using SUSI in autocollimation from the beam reducing telescope. A tungsten lamp was used as a source, and various numbers of frames were integrated. In these cases there was no attempt to remap the data for dispersion, and the data were processed from the detector coordinates. An artifact at a spatial frequency of 96 can be seen, and also a much smaller one at 64. Each of the data sets has been scaled down by increasing powers of 2, so that they all appear to have the same total power.

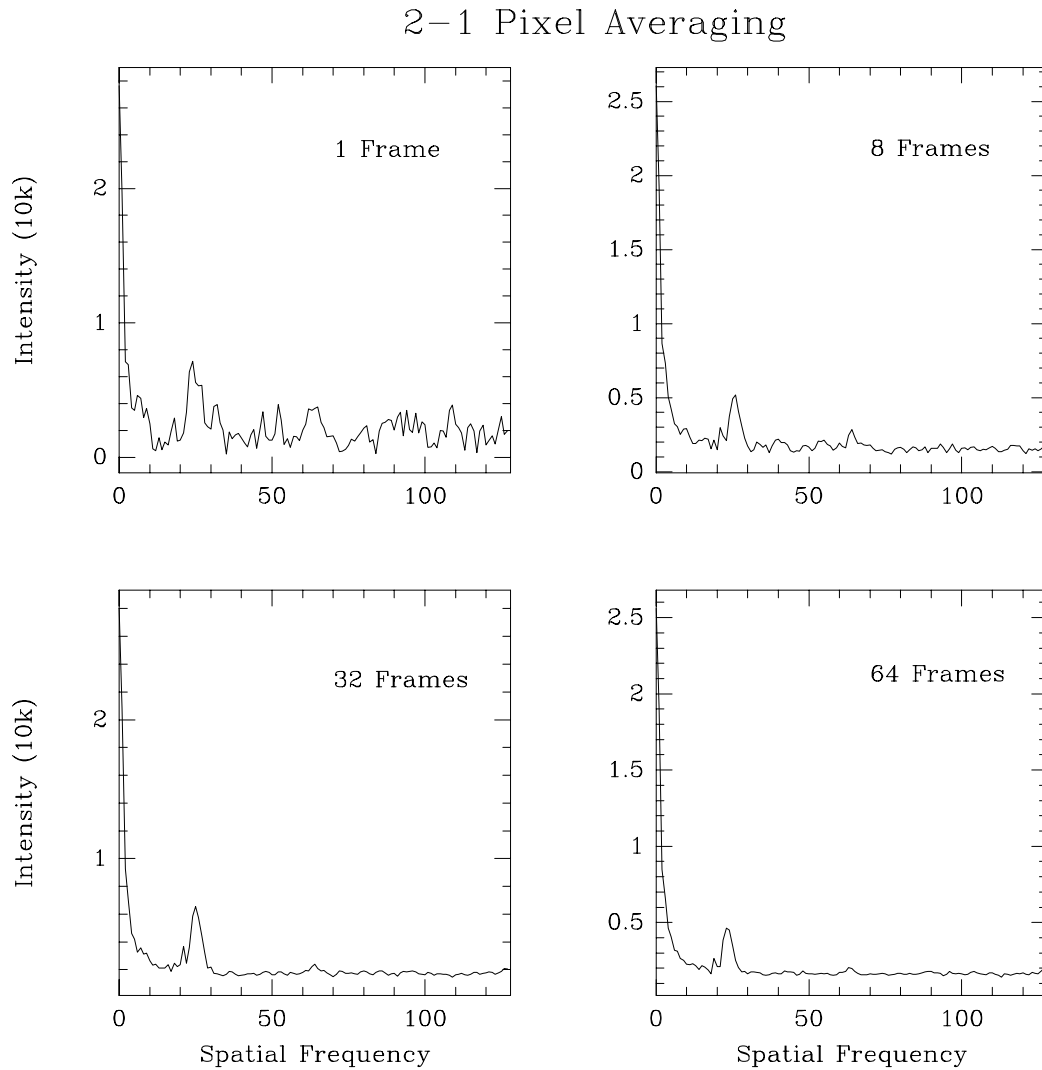


Figure 10.10: Here, data from neighboring pixels have been averaged together. There are then 128 pixels of information padded with 128 zeros. These observations were done using SUSI in autocollimation, using a tungsten lamp as an artificial star. The artifact at a spatial frequency of 96 is now absent, although the one at 64 is still visible.

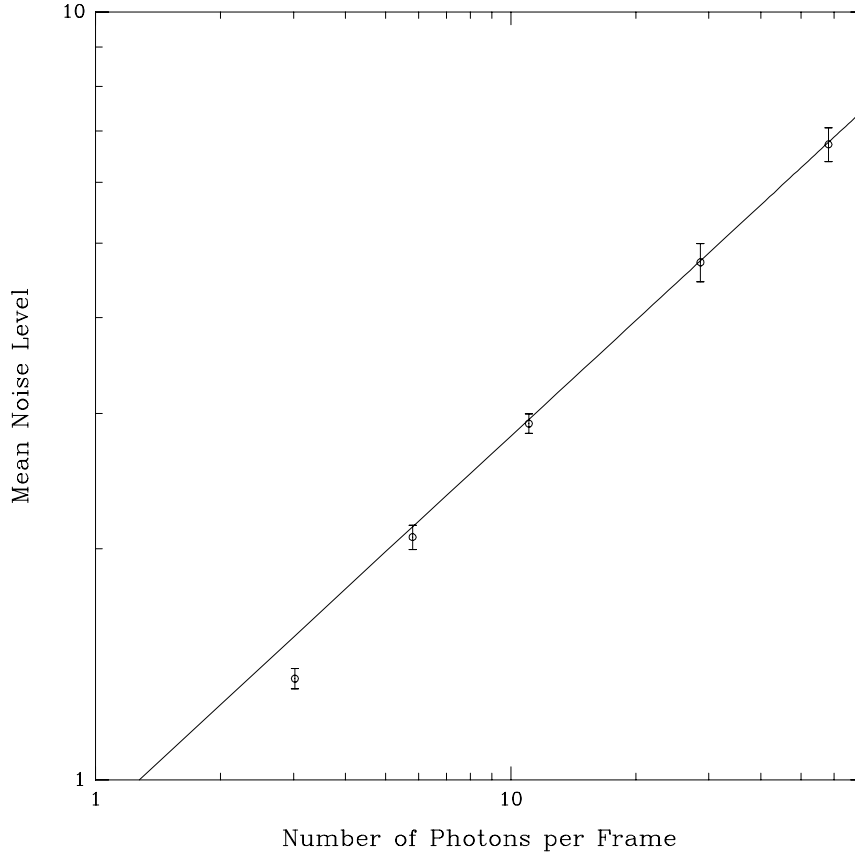


Figure 10.11: The mean noise level as a function of the number of photons per frame. The noise level in DFT amplitude estimates is predicted according to the Rayleigh distribution, as discussed in §7.3. The straight line is the theoretical model, $\langle Z \rangle = (\pi\sigma^2/2)^{1/2}$, where $2\sigma^2 = N_t$ the number of photons-per-frame. This agrees well with the measurements down to light levels below 10 photons-per-frame.

The mean noise level of amplitude estimates

Measurements were also made of the mean noise level as a function of the number of photons-per-frame. The data processor calculates DFT amplitudes from the channeled spectrum, and according to the theory outlined in §7.3 the noise should follow a Rayleigh distribution.

SUSI was therefore set up in autocollimation, but the path compensating carriage was positioned so that no fringes would be present. A tungsten lamp was used with neutral density filters, and the frame time was varied to reduce the number of photons-per-frame. The results are shown in Fig. 10.11. The solid line represents the theoretical model, $\langle Z \rangle = (\pi\sigma^2/2)^{1/2}$, where $2\sigma^2 = N_t$ is the number of photons-per-frame. This is the mean value of the Rician distribution when $\rho = 0$, shown in

Eq. 7.12. The data from the measurements is indicated in the figure. It can be seen that even down to levels of less than 10 photons-per-frame, the measurements agree remarkably well with the theory.

10.3 Conclusion

Figure 10.7 showed the response of the group delay processor to fringes of increasing path-difference. Under ideal conditions the peak would be 0.50 of the DC level. This was shown previously by the model of Eq. 5.1 and the illustrations in Fig. 5.2. However, in the experiments the peak was observed to be closer to 0.15 of the DC level. The difference is due to the assumptions made in Eq. 5.1.

- The spectrum of the tungsten lamp and the wavelength sensitivity of the camera results in a non-uniform envelope of the channeled spectrum. The Fourier transform of this envelope is convolved with the fringe peak and therefore reduces its height.
- The fringes were mapped with a non-constant dispersion across the detector. It was found that re-mapping to correct for dispersion did not significantly improve the sensitivity, despite the predictions outlined in §6.2. In practice, a straight averaging to 128 pixels produced better tracking. This removed the camera artifact at a spatial frequency of 96 and with zero padding improved the delay sampling to intervals of $1.0 \mu\text{m}$.

Moreover, the visibility of the fringes will always be less than 1.0, due to diffraction effects, slight imperfections in the optics, and seeing within the enclosure.

The difference between the actual and ideal response can therefore be taken into account when predicting the limits of performance, now using a visibility amplitude of 30% or .15/.50 in the equation for the probability of tracking loss, Eq. 7.20. This implies that active tracking would probably not be possible at light levels lower than 10 photons per frame, because more than 1000 frames would be necessary to guarantee tracking. If the frame times are several milliseconds long then the integration time required is several seconds for a stationary delay.

Chapter 11

Stellar Observations

In this chapter observations of α CMa, α PsA, and α Eri are presented, and tests of the delay tracking servo are shown. These tests show the first close-loop operation of a fringe tracking servo based on group delay tracking.

The path-difference detected by delay tracking contains two components: a group delay caused by the atmosphere, and path-differences caused by tracking errors of the carriage. The carriage errors are measured independently by the laser metrology system at SUSI, and are illustrated for comparison with the fringe tracking.

11.1 Observations of α CMa

α CMa was observed in April 1993. The changing path difference was easily detected with the camera and processor, but at that stage in the commissioning of the path compensator the path motions were dominated by tracking errors of the carriage.

The path compensator prior to May 1993

All observations taken prior to May 1993 suffered large excursions in the path correction. Only the speed of the carriage motors was adjusted in the servo with the laser metrology system, and the piezos at the catseyes had not yet been implemented. The path errors were a function of carriage speed and would exceed amplitudes of $50\ \mu\text{m}$ at the higher rates. Some examples are illustrated in Fig. 11.1. These were typical of the path errors that existed when α CMa was observed.

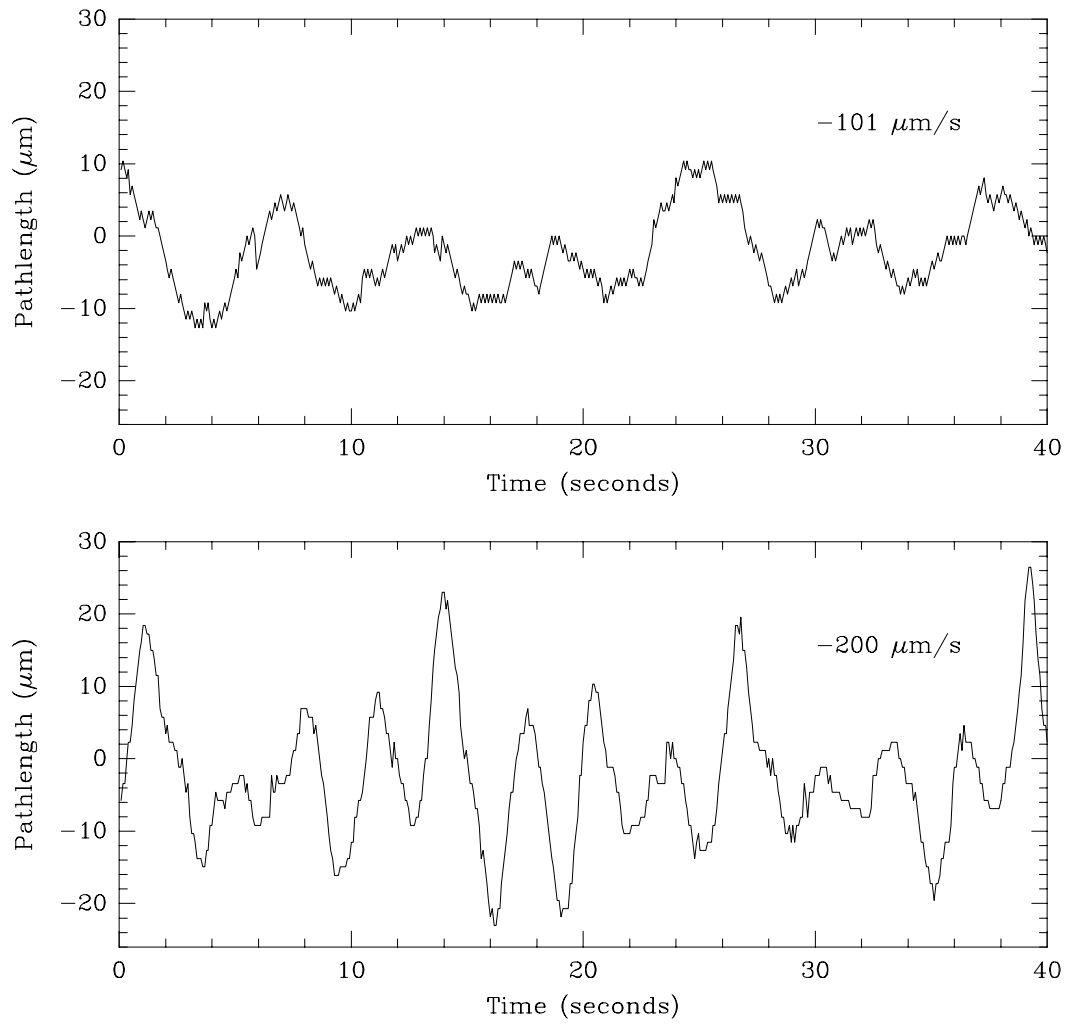


Figure 11.1: The position errors of the optical pathlength compensator as a function of the speed of the path compensating carriage. These measurements were made by the metrology system. The plots are for carriage velocities of $-101 \mu\text{m/s}$ and $-200 \mu\text{m/s}$.

The observations

α CMa was observed on two nights. A 5 m baseline was used, the apertures were 30 mm in diameter, and the seeing conditions during the observations were between 1.0 and 1.3 arcseconds. Frame times of 2.8 ms produced data rates of 50–55 photons-per-frame, and integrations of 20 to 32 frames were used (values of \mathbf{a} between 243 and 248). The data were averaged to 128 pixels and padded with 128 zeros before being Fast Fourier Transformed.

The location of the peak in the power spectrum was recorded every fifth frame for up to 30 seconds and was displayed in real-time on an oscilloscope. Records of the carriage motions were made concurrently and the two data sets are shown overlaid in Figs. 11.2, 11.3, and 11.4. The spectrum of α CMa that was detected by the PAPA camera is shown in Fig. 11.5.

Each of the small crosses in Figs. 11.2–11.4 represent the location of the peak in the power spectrum. In each case the dominant component of the pathlength change is due to the motions of the carriage, with residual differences on a scale of $5\ \mu\text{m}$ rms: these differences are what would be expected from the effects of atmospheric turbulence on a 5 m baseline. In the upper graph of Fig. 11.2 there are occasions where the number of fringes drops below the lower threshold and there appear ‘plumes’ of tracking loss: this can be seen at 0 and 4 seconds. The signal to noise ratio in these plots is not high. Sample power spectra, taken at times when fringes were detected, are shown in Fig. 11.6. In each case the location of the fringe is easy to identify, but is present on a broad pedestal of noise.

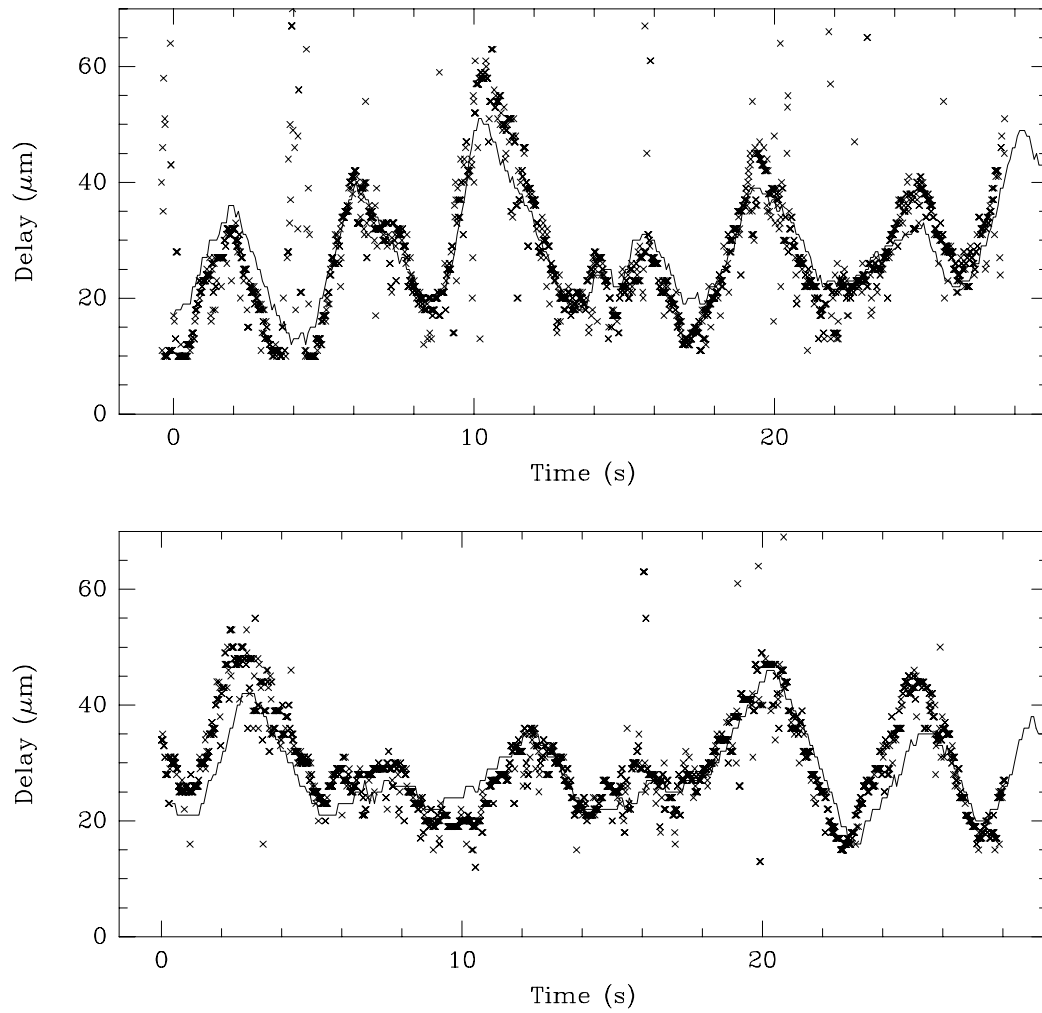


Figure 11.2: The path-difference changes observed with α CMA on April 3, 1993. The solid line in each graph represents the position of the carriage as measured by the laser metrology system. The small crosses indicate the location of the peak in the power spectrum of the fringes. There are 2000 peaks recorded in each plot.

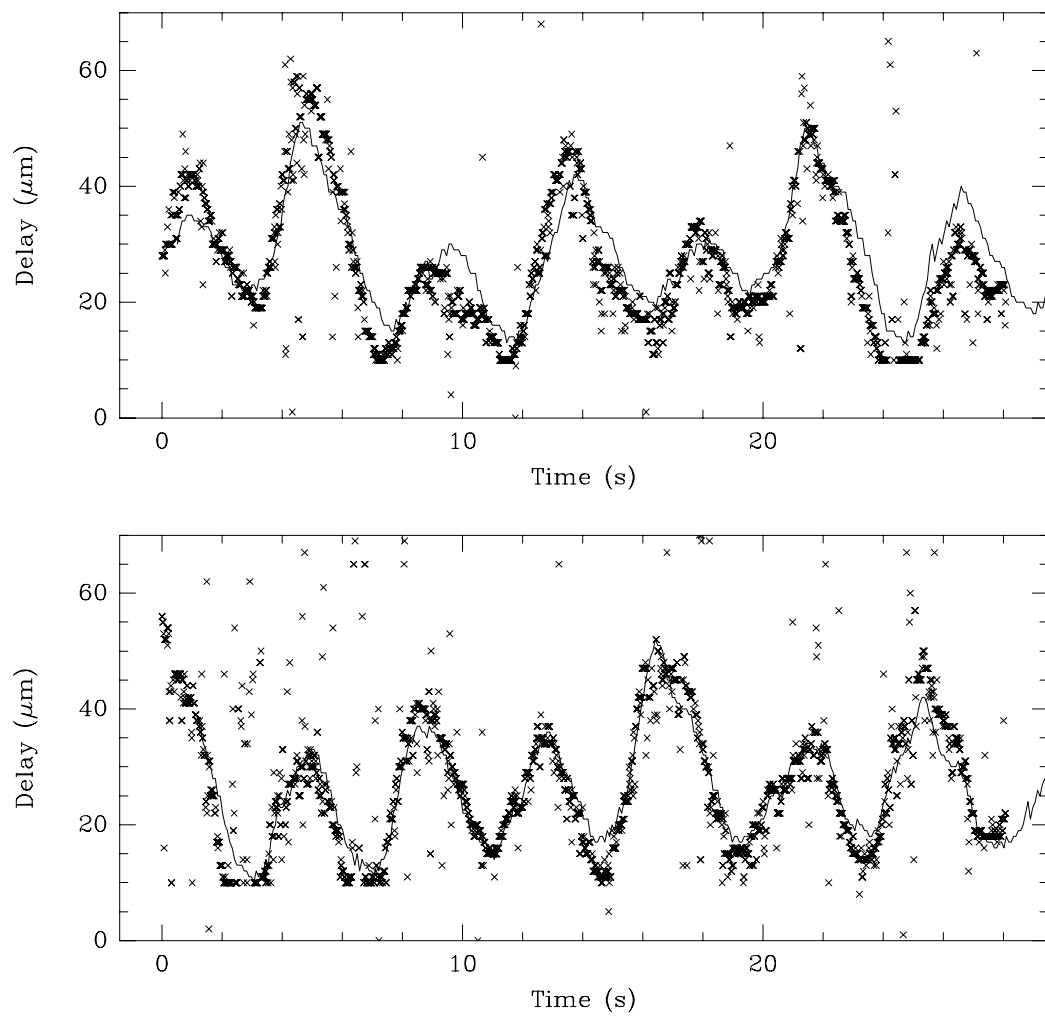


Figure 11.3: Further traces of the path-difference changes observed with α CMa on April 3, 1993.

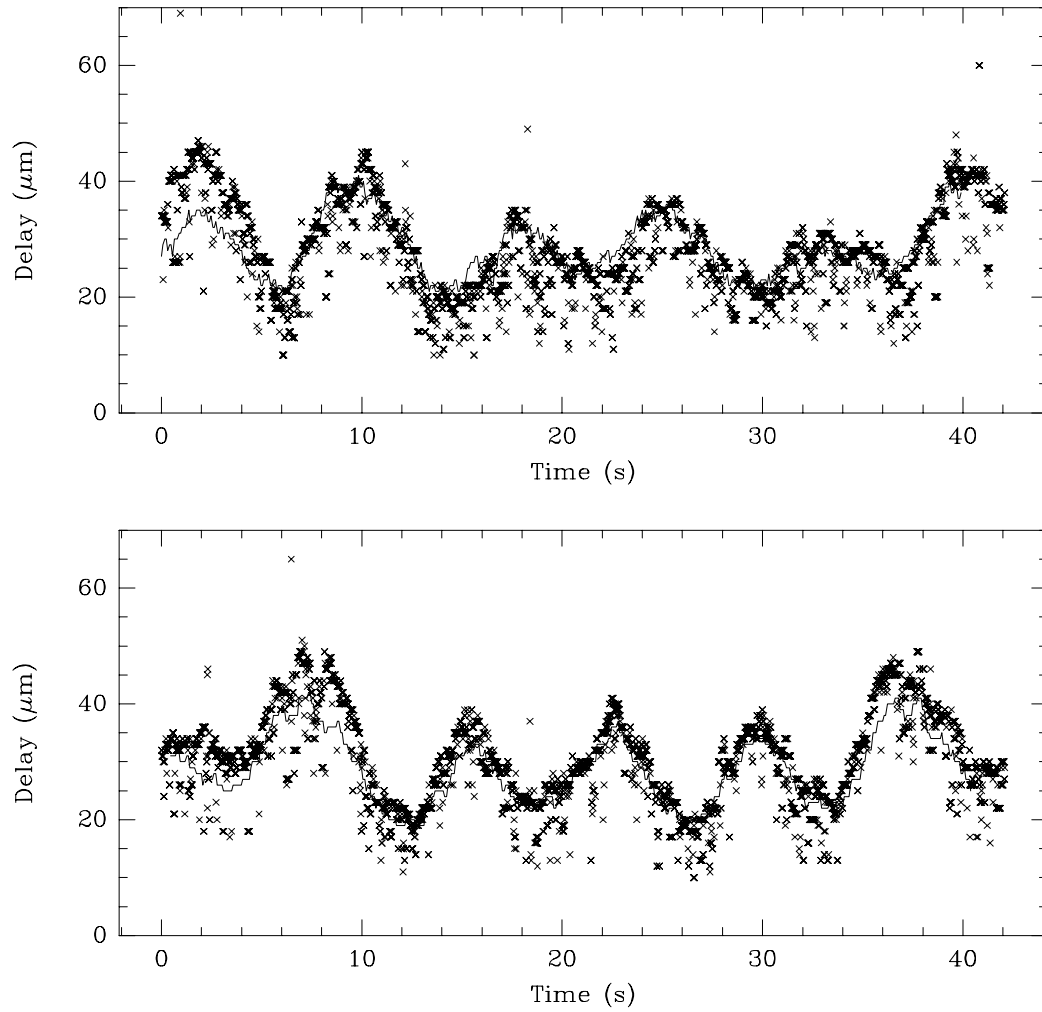


Figure 11.4: Path-differences observed with α CMA on April 4, 1993.

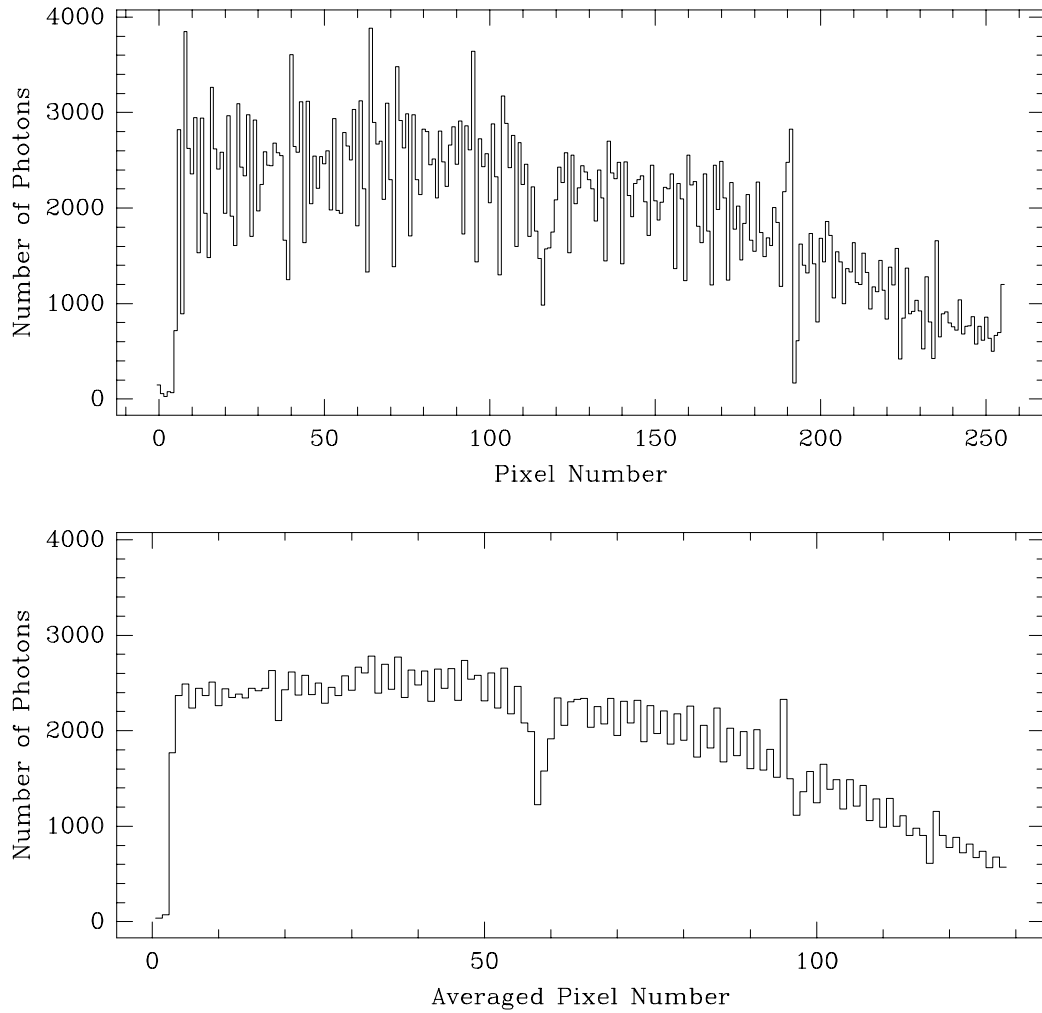


Figure 11.5: The spectrum of α CMa obtained as a cross section of the previous PAPA camera image. The top spectrum shows the complete 256 pixels. The bottom shows the effect of averaging neighboring pixels together. For all observations of α CMa the data were averaged in this way as it arrived. The H β line at 486.1 nm is visible as an absorption feature near the center of the spectrum.

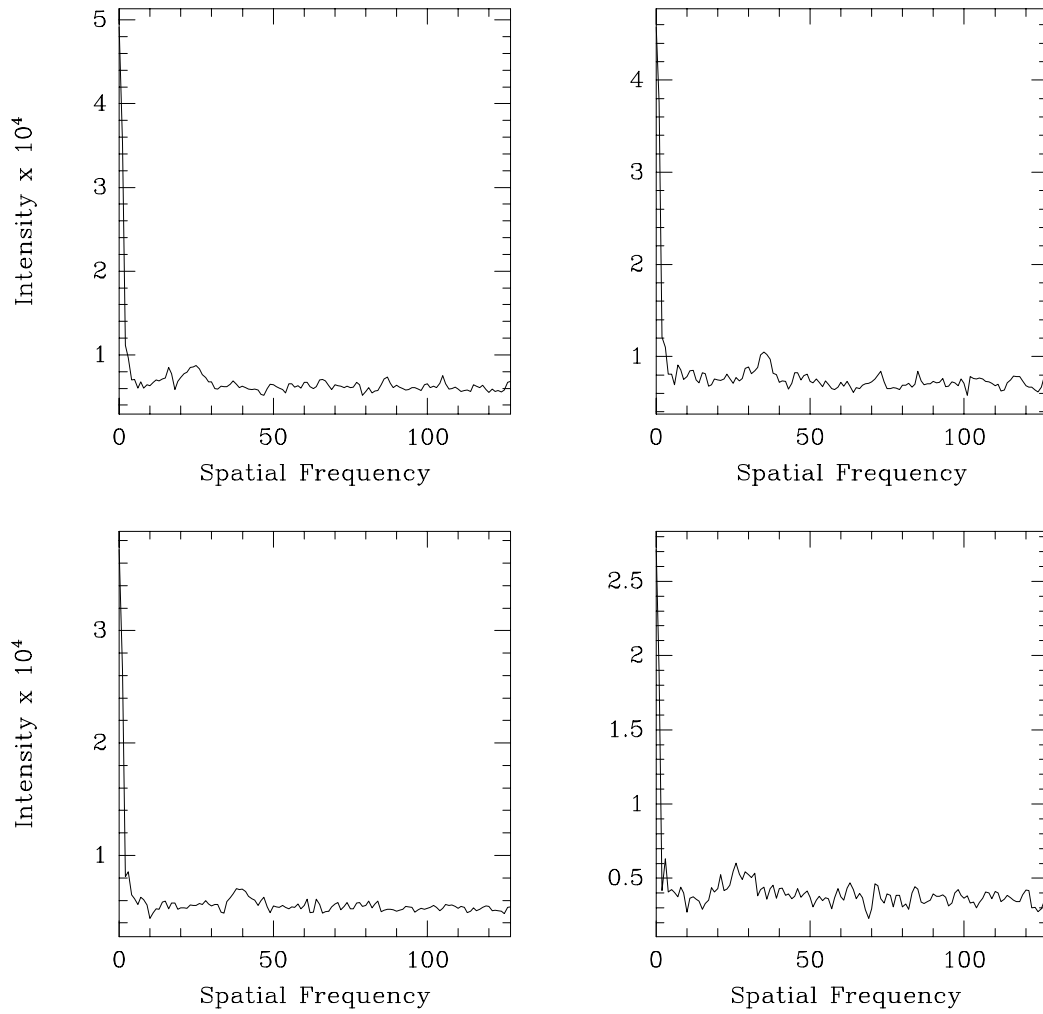


Figure 11.6: These plots are representative of the spatial frequency spectra which were used to track the group delay that was illustrated in Figs. 11.2–11.4.

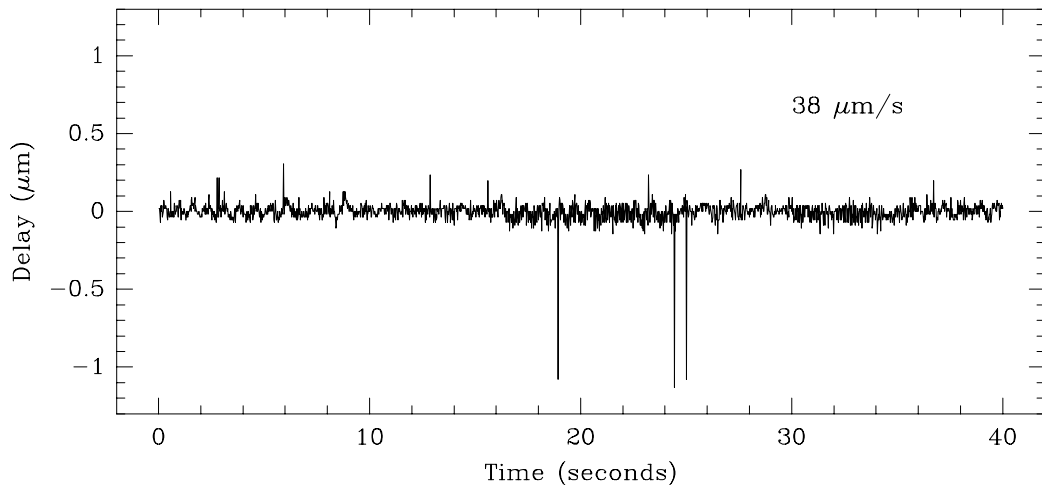


Figure 11.7: The position errors of the optical pathlength compensator in September 1993, after changes to the laser metrology system. These measurements were made by the metrology system only. The plot is for a carriage velocity of $38 \mu\text{m/s}$ which was the highest velocity used for observations of α PsA and α Eri. The carriage speed is slower than for the observations of α CMa, because these stars were closer to transit when observed.

11.2 Observations of α PsA and α Eri

The stars α PsA and α Eri were observed in early September 1993, and since April the carriage and metrology servo had been markedly improved; atmospheric effects now dominated the path variations.

The path compensator in September 1993

The performance of the path compensator was improved through the work of Andrew Booth and Stephen Owens. Between May and August the electronics for both metrology counters were rebuilt, the clock for the metrology servo was sped up to a 2 ms period, and the piezo mirrors on the carriage were included in the servo loop. This allowed the path compensator and metrology system to servo with rms path errors of only $0.1 \mu\text{m}$. The carriage errors that existed in September are shown in Fig. 11.7. The $\sim 1 \mu\text{m}$ spikes are not real, but are generated by bit errors in the metrology electronics. Because the delay tracking had a resolution of $2 \mu\text{m}$, any changes that could be resolved could not be caused by the carriage, and the remaining changes were therefore due to the atmosphere.

The observations

Each observation used a 5 m baseline and aperture diameters of 60 mm, under conditions where the seeing varied between 1.5 and 1.7 arcseconds. A bandwidth of 110 nm, centered on 495 nm, was imaged on the PAPA camera and its 256 pixels were averaged to 128. The resolution in delay was therefore $1/\Delta\kappa = 2.10 \mu\text{m}$, and zero padding with 128 zeros allowed the delay to be sampled at intervals of $1.05 \mu\text{m}$.

Observations of α PsA

The observations of α PsA are shown in Figs. 11.8 and 11.9. As before, the power spectra of the channeled fringes were integrated recursively, and the maximum in the integrated spectrum was recorded after every fifth frame (every 20 ms when using a 4 ms frame time). The maxima are indicated by the small crosses in each graph and 1500 points are shown for 30 seconds of data. With α PsA, frame times of 4 ms produced 22–26 photons-per-frame.

The motion of the delay can be traced through each plot. The peak-to-peak changes are between 12 and 18 μm , and occur over timescales of approximately 8 seconds. Short term fluctuations of 5 μm per second are common, and are occasionally sustained for two or three seconds.

Observations of α Eri

For the observations of α Eri frame times of 4 ms were again used and an average of 50 photons-per-frame were observed. Traces of the fringe motion are shown in Figs. 11.10, 11.11, and 11.12. In each case the observed fringe motions were similar to those seen with α PsA, but the signal-to-noise ratio is now improved because of the higher count rate. This can be seen by comparing Fig. 11.10 with the previous plots.

In Figs. 11.11 and 11.12 the fringe motions have been filtered to remove the noise. The upper plots in each figure show the raw data that was detected by the Max-find module in the processor. The traces are now drawn so that all data points are connected by lines: this better illustrates the effect of the filtering. The tall spikes are due to the noise peaks which appeared as outliers in the previous traces. The lower plots are the same data after filtering by the software. Not only has a pre-filter been used to remove the noise events, but the data have been smoothed as well. It can be seen by comparing the upper and lower plots that the motions of the fringe have been faithfully reproduced. The discontinuities in the lower plots show instances where the tracking was lost; a 0 was then recorded to signal that no information was available.

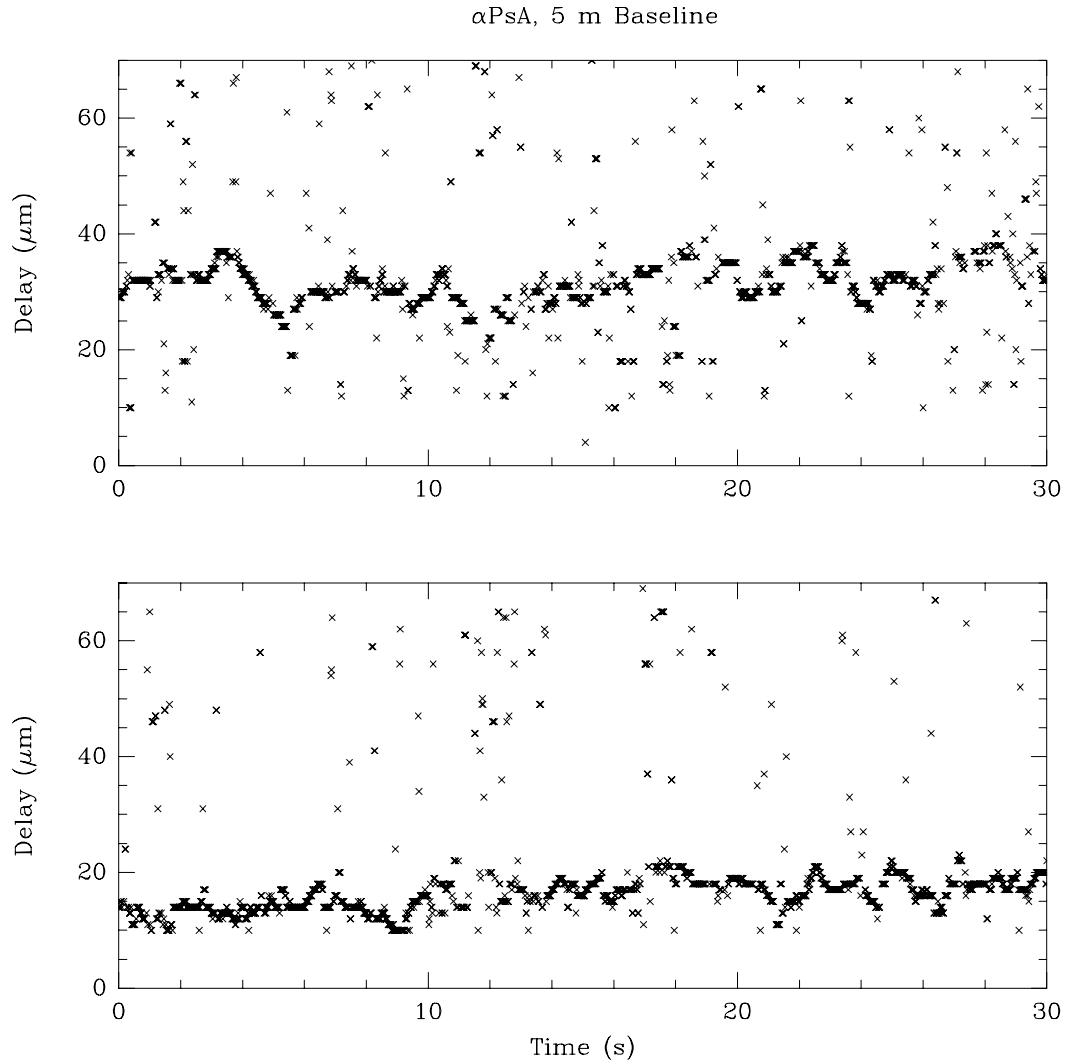


Figure 11.8: The group delay wander observed with α PsA on Sept. 2, 1993. The metrology errors were only a fraction of a micron and did not contribute to the path motions visible in these graphs. These are traces of fringe motion due to atmospheric turbulence. The small crosses indicate the location of the peak in the power spectrum of the fringes. An aperture diameter of 60 mm was used with frame times of 4 ms and seeing conditions of 1.5 arcseconds. The recursion used a filter constant of $a=248$, equivalent to summing 32 frames and required 3 FFT scalars. The carriage speed was $16 \mu\text{m/s}$.

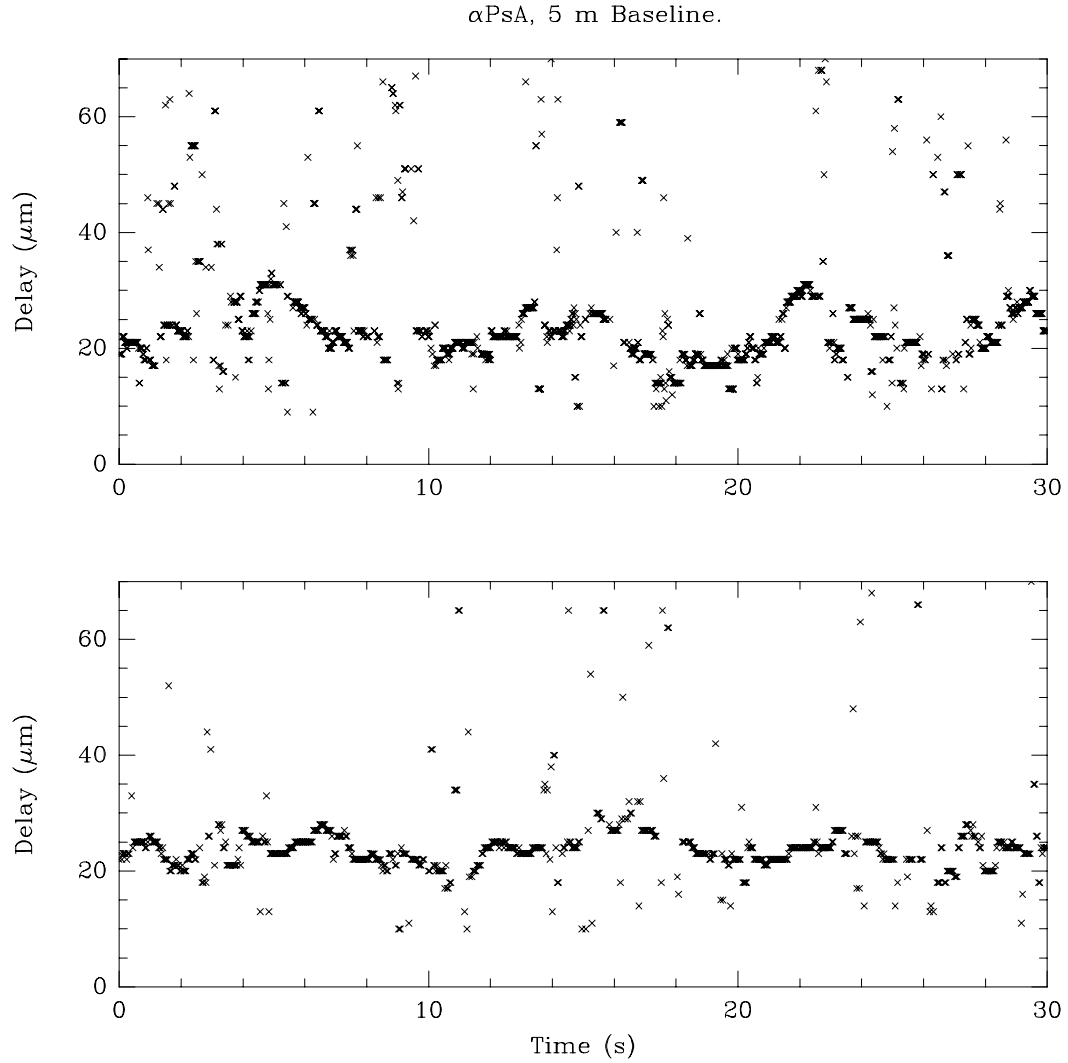


Figure 11.9: Group delay wander observed with α PsA on Sept. 3, 1993. These are traces of fringe motion due to atmospheric turbulence. The upper figure used a recursion equivalent to summing 51 frames ($a=251$) with a frame time of 4 ms. The lower figure used 28 frames ($a=247$) and a frame time of 8 ms. In each case 3 FFT scalars were used. The speed of the carriage was $17 \mu\text{m/s}$, and the aperture size was again 60 mm in diameter. The seeing was 1.2–1.5 arcseconds.

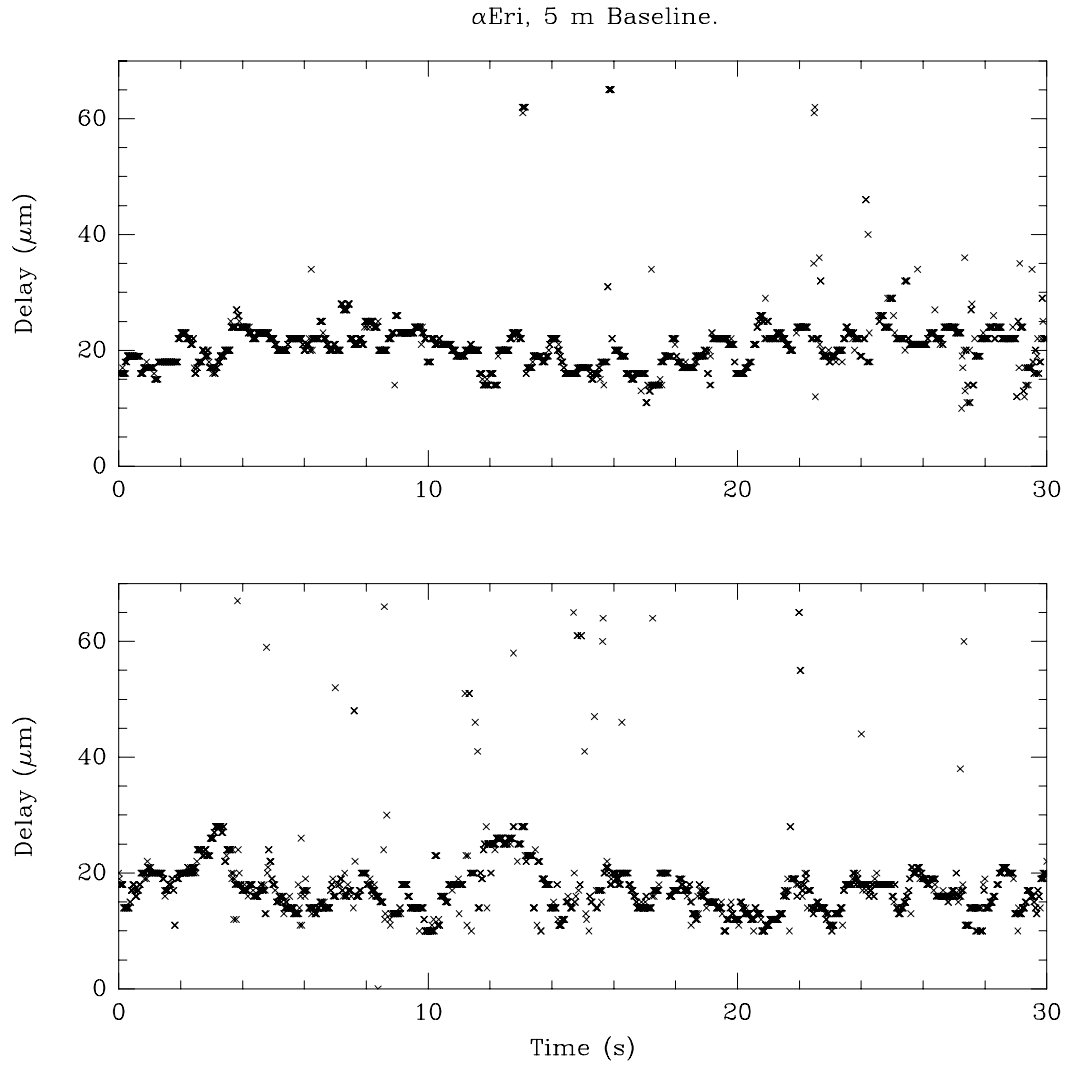


Figure 11.10: Fringe motions observed with α Eri on Sept. 2 and 3, 1993. In both cases a 60 mm aperture was used with a frame time of 4 ms, 4 FFT scalars were enabled, and the seeing conditions were 1.5–1.6 arcseconds. The upper figure used 64 frames ($a=252$) and the lower used 32 ($a=248$). The trace of the carriage motions, shown in Fig. 11.7, were taken at the same time as record shown in the lower figure.

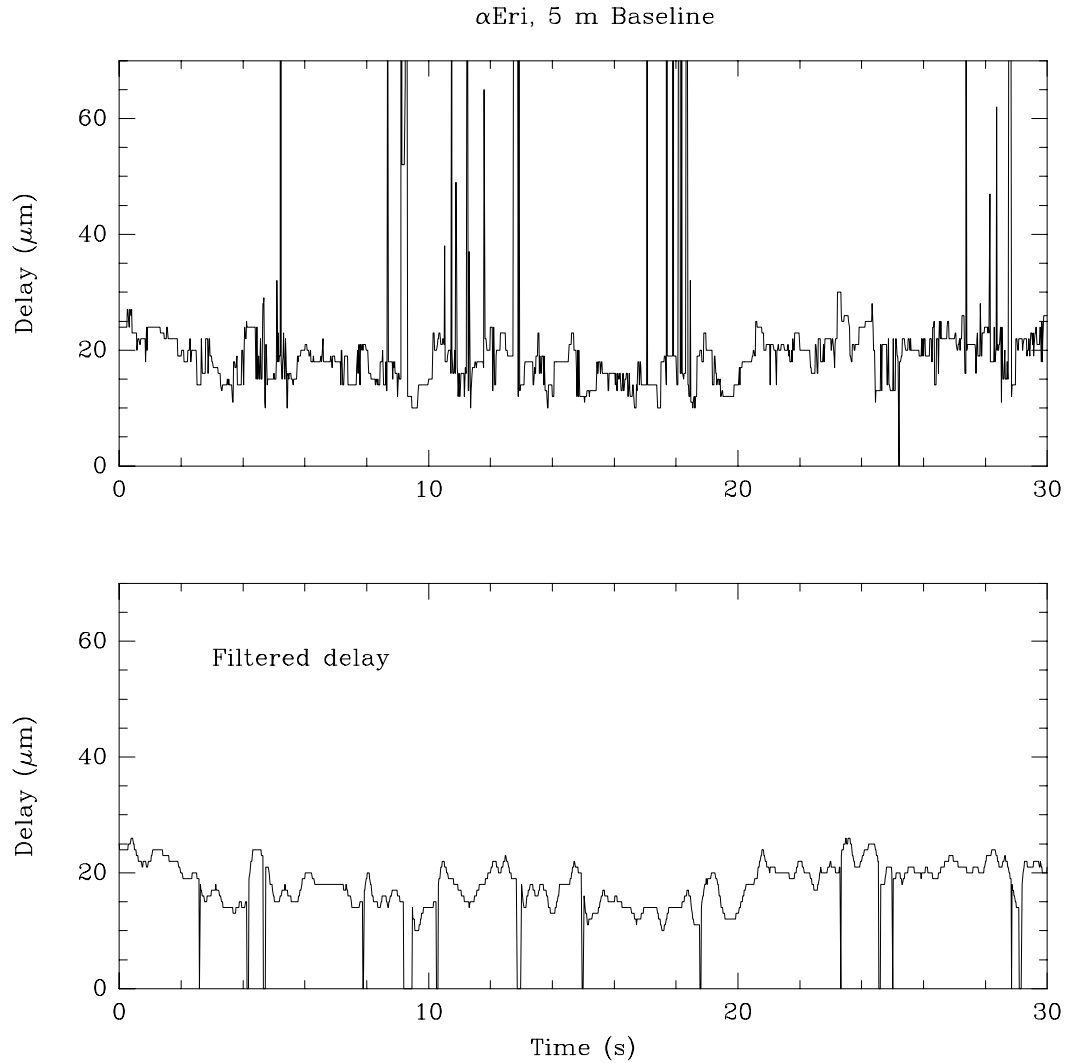


Figure 11.11: Filtered fringe motions observed with α Eri on Sept. 4, 1993. The upper figure shows the raw data as processed by the hardware's 'Max-find' module. The scattered noise that appeared in previous records now appears as sharp spikes since the data points have been drawn here with connecting lines. The lower figure shows the same data after the spikes have been removed and the delay filtered, as discussed in Chapt. 9. The discontinuities in the lower graph are intentional: a delay of 0 is sent to the carriage servo when the tracking is lost. 64 frames were integrated ($a=252$) with a frame time of 4 ms and 4 FFT scalars were used. The seeing was 1.6 arcseconds. An aperture diameter of 60 mm was used.

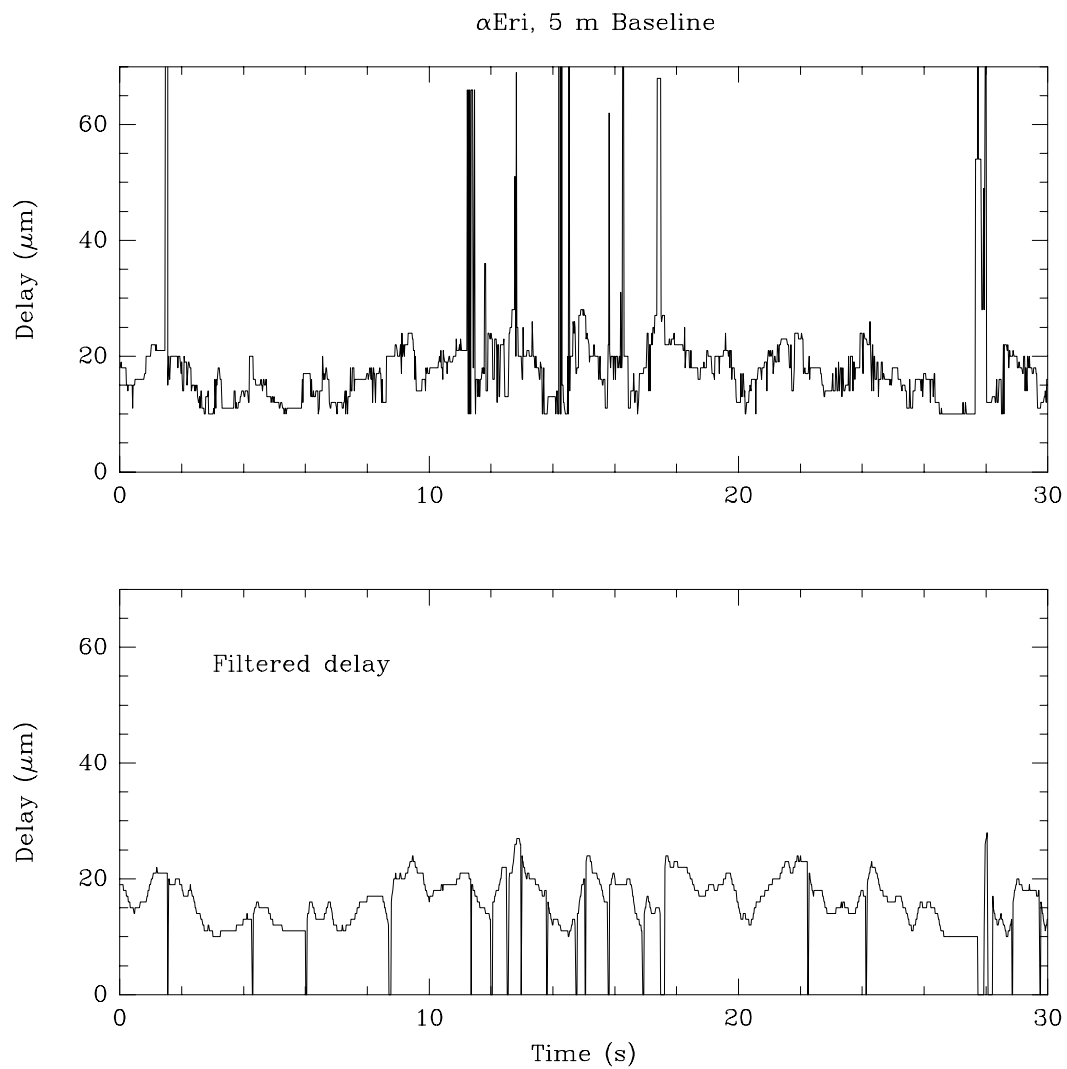


Figure 11.12: Filtered fringe motions observed with α Eri on Sept. 4, 1993. This trace was observed under conditions that were the same as with the previous example.

11.3 The Path Compensating Servo

Preliminary tests with the fringe tracking servo were carried out on September 7, while observing α Eri. The servo was designed to track in an offset-mode, with between five and ten fringes across the detector. This is because when there are a small number of fringes their detected visibility is higher, although when there are less than five the tracking signal becomes lost in the low frequency structure of the stellar spectrum. At zero pathlength the fringes disappear entirely.

The servo equations were added to the program `carservo` by Andrew Booth. This program is responsible for the transfer of data between the computers that determine a star's position, generate rates for the path compensating carriage, and monitor the feedback from the laser metrology system. The test servo was a proportional integrating controller: the correction was proportional to the input, was responsive to long term drifts, but did not contain a damping term. The group delay estimates were read by `carservo` every 80 ms, which was adequate because resolvable path changes would take place over hundreds of milliseconds.

Two servo equations were required: a positive and a negative servo. The number of fringes in the channeled spectrum only determines the distance from zero path-difference, but not whether the error is positive or negative. It cannot indicate whether the observations are taking place before or after the peak in the coherence envelope. If the wrong servo were chosen then small errors should cause the carriage to drive *away* from the offset position. In practice this was found to be true in every case, but its results were different depending on the initial conditions. There are two cases to consider:

1. If the path-difference was above the tracking offset, then turning on the wrong servo would simply increase the path-differences and drive the carriage further out on the coherence envelope. In these cases the number of fringes would increase monotonically, and the tracking would indeed be lost.
2. If the path-difference was less than the tracking offset, then turning on the wrong servo would still drive it away from the offset position, but it would drive it towards zero path-difference, and then over the top of the coherence envelope. Once the path-difference went through zero and changed sign, then the sense of the servo would be correct. The servo would search for the correct offset and begin tracking there.

Numerous time during the tests, it was observed that if the sense of the servo was changed, that the servo would travel over the top of the delay curve and lock onto the offset at the other side.

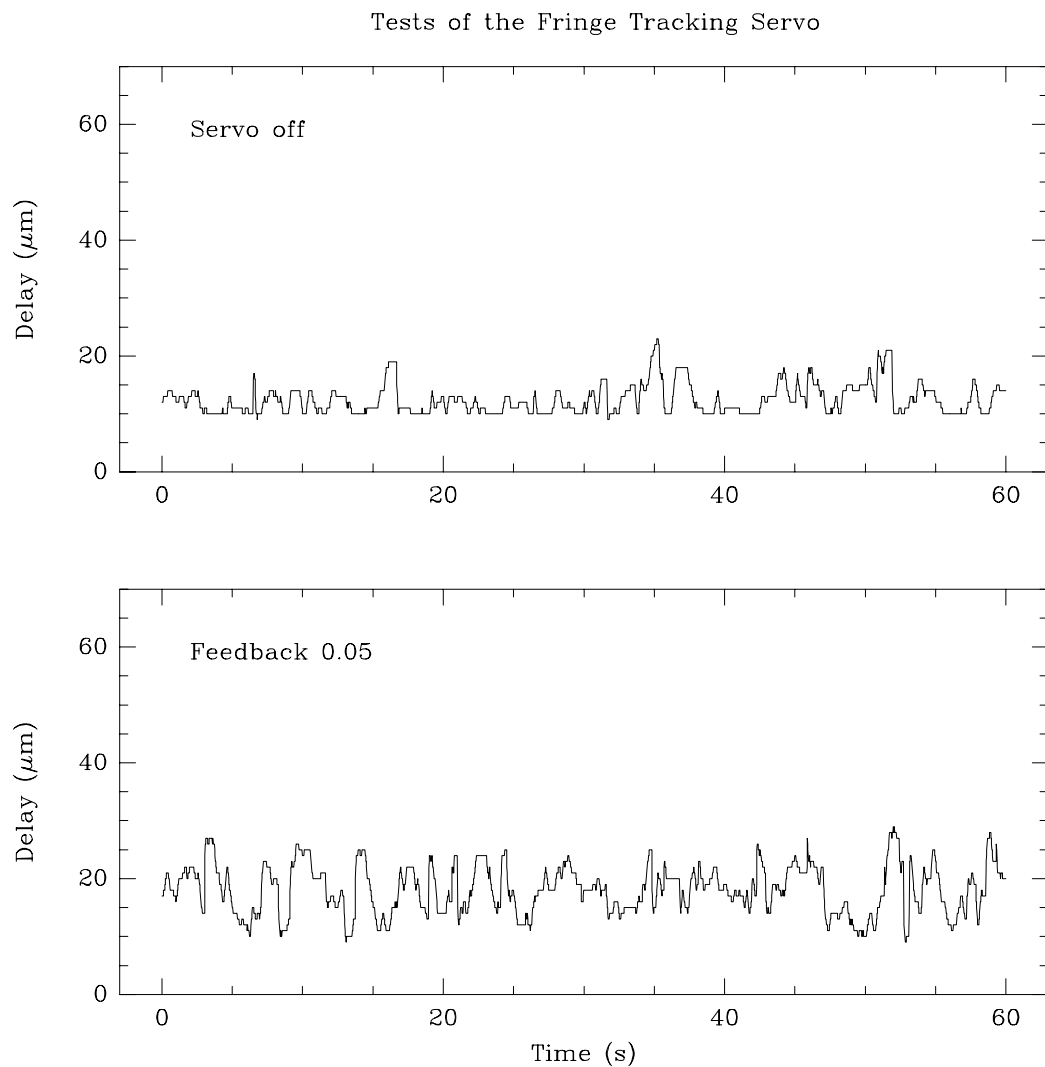


Figure 11.13: Test of the fringe tracking servo with α Eri on Sept. 7, 1993. The location of the delay was read by the servo which controls the path compensating carriage. For these tests there was no damping in the servo equation; only the feedback constant was adjusted. Shown here and on the following page are records of the pathlength change as a function of the feedback constant. It is obvious from these that although the servo functions, the feedback constant is too high in every case.

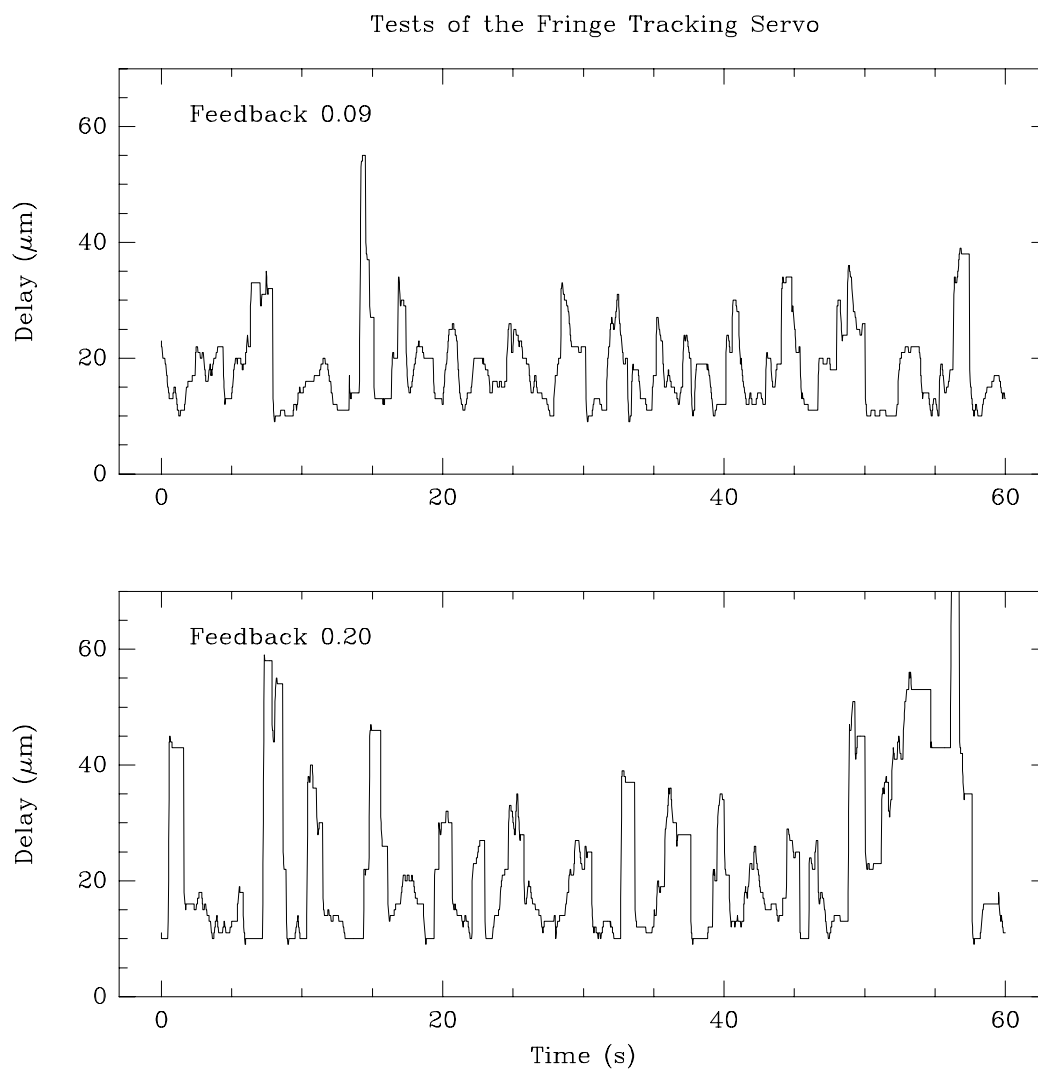


Figure 11.14: Test of the fringe tracking servo with α Eri on Sept. 7, 1993. As the feedback constant is increased the amplitude of the pathlength change increases. Further tests are necessary to properly tune the servo.

The tests of the servo are shown in Figs. 11.13 and 11.14. The response of the group delay tracking is shown for increasing values of the feedback parameter. It is obvious from these figures that the servo had locked onto the fringe, but the feedback parameter was too high in every case, and closing the servo loop caused the path-difference to oscillate. Unfortunately, tuning it required more time than was available, and poor weather brought the observations to a close.

11.4 Conclusion

The emphasis of both the laboratory experiments and the stellar observations has been to evaluate the performance of active fringe tracking. The results indicate that for this to be successful, the observations typically require 20 or more photons-per-frame and a recursive integration of 32 frames. With half as much light, it is still possible to locate the delay in moments of good seeing, but the probability of tracking failure is much greater.

However, at higher light levels the changes in atmospheric phase-difference could be identified, and the system was observed to work well. The motions observed with α CMa closely followed the large path variations of the carriage, as can be seen in the overlaid records taken with the metrology system. The fringe tracking servo was closed for the first time, using observations of α Eri. Although its feedback parameter remains to be tuned, the servo was observed to be robust and stable. Its behavior also indicated that if the servo were started near the peak of the coherence envelope that it would locate the proper tracking offset, whether the sense of the servo was positive or negative.

The atmospheric phase changes observed with a 5 m baseline were consistent with the predictions given in Chapt. 8. The peak-to-peak motions were 12–18 μm , and changes of 5 $\mu\text{m/s}$ were common over timescales of 1 or 2 seconds. These features are what would be expected, and are also evident in the fringe motions presented by Colavita et al. (1987), seen while observing α Aql with a 12 m baseline.

Chapter 12

Conclusions

This thesis has been devoted to the development of a new fringe tracking system, based on group delay tracking with the Fast Fourier Transform. The work was divided into four parts, beginning with an introduction to fringe tracking and a review of previous work.

Part II described the PAPA camera that was commissioned for use in this project. It was based on a model that had been constructed at Harvard University, but which proved to have numerous design faults. These gave rise to image artifacts that could not be corrected without redesigning the camera's optics. The causes of image artifacts were investigated: the descriptions of lens tolerances, alignment errors, image folding, and the sum-of-all-masks represent original contributions by the author. Based on this work, the design of a new camera was presented. It incorporated a new mask-plate that was built to correct for vignetting artifacts, and a method of lens mounting which obviated the need for tilt-plates. The success of the new detector was then demonstrated.

Part III described the theory of group delay tracking and discussed its low light level performance. The principal result was a derivation of the probability of tracking loss. A figure of merit was also derived for the minimum number of photons required to guarantee tracking. The fundamental limit of operation was shown to be reached when there is at most 1 photon-per-frame: the spatial power spectrum of the fringes is then featureless, and no information can be extracted. At higher light levels the success of tracking depends on the visibility of the fringes, the resolution in delay, the allowable coherence loss, and the speed of the path-changes. For moving fringes it was shown that if the delay was changing at a constant rate, and a 110 nm bandwidth were used then at least 9.7 photons-per-frame would be needed to guarantee tracking.

In Part IV the experiments were described. In the laboratory it was found that sinusoidal pathlength changes were detectable at 10 photons-per-frame, providing the amplitude was less than 20 μm and the period less than 2.5 Hz. At SUSI it was found

that light levels of 20–50 photons-per-frame were required for active tracking during observations of α PsA and α Eri. The fringe tracking servo was then demonstrated with observations of α Eri. The observed limitations of group delay tracking are in accord both with the theory derived in Part III and with prior simulations of Buscher (1989), as noted in Chapt. 2.

There remain several improvements and extensions of this research which are worth discussing. Therefore, in conclusion some suggestions for future work are given.

12.1 The PAPA Camera

The Detective Quantum Efficiency (DQE) of the camera could be improved using a different image intensifier package. There are two well defined problems that need to be addressed:

- The microchannel plate in the Varo 3603 tube has an open-area ratio of 57%, and therefore blocks almost half of all photoelectrons that are generated at its cathode. The detective quantum efficiency of the tube is reduced accordingly.
- The pulse-height distribution of photon events is a negative exponential, and it is impossible to separate photon events from noise events. This is because the microchannel plate is biased for night-vision operation rather than for photon counting.

These limitations are common with other array detectors which use microchannel plates for photon counting. There is active research in this field and it is likely that a suitable solution will be found. Harvard had suggested using a hybrid combination of intensifiers as an alternative. Although that approach may yet prove successful, testing it was beyond the scope of this work. By this method, it is claimed that the detective quantum efficiency could be increased by a factor of four, to 8 or 10%.

An improved DQE would extend the number of stars that could be actively tracked with this system. The number of photons-per-frame that are detected by the PAPA camera is shown in Fig. 12.1, where it is assumed that 120 nm of bandwidth is used, from 430–550 nm, but with different aperture sizes and integration times. The measurements of α PsA and α Eri were conducted with a 4 ms sample time and 60 mm diameter apertures, and their magnitudes are indicated in the figure. A factor of four improvement in sensitivity would mean that the same number of photons-per-frame would be observed at a magnitude of 2.9 as were observed with α PsA.

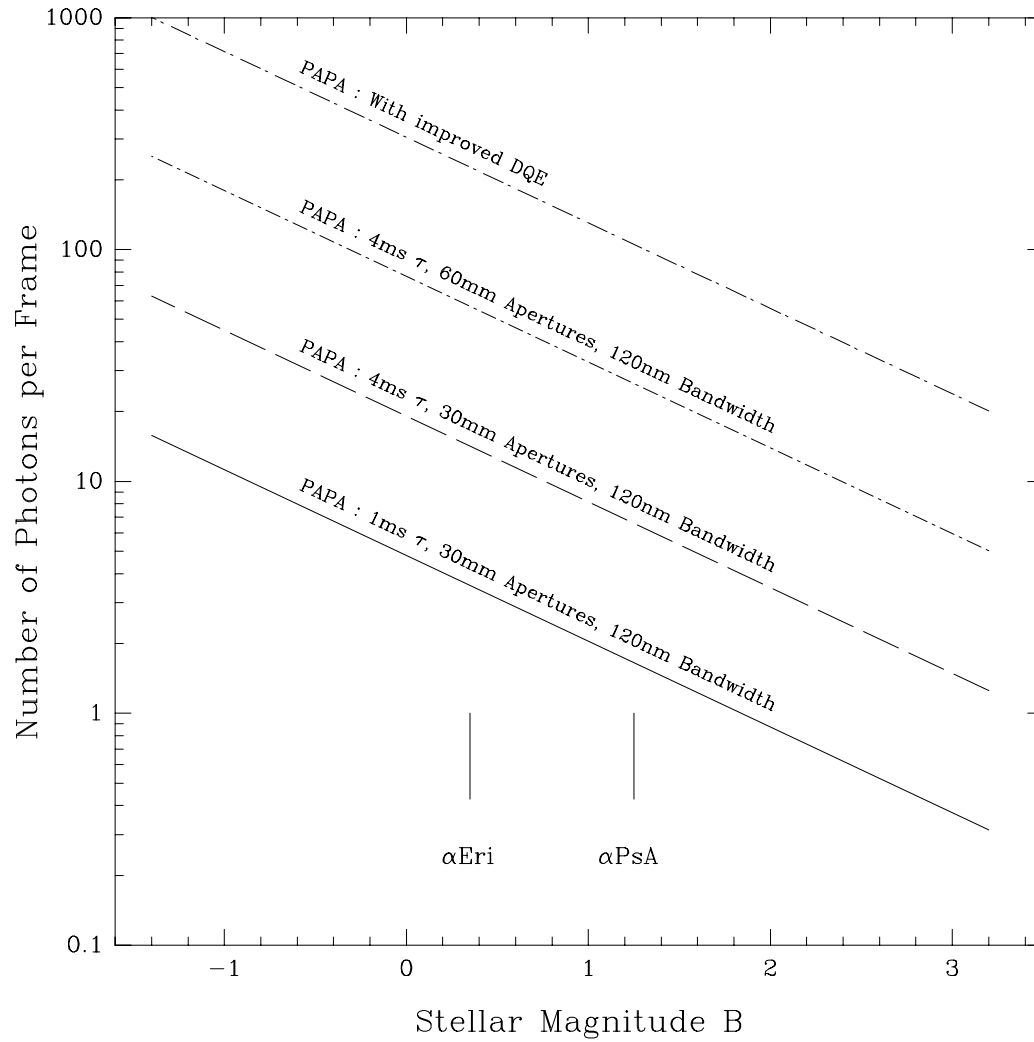


Figure 12.1: The number of photons-per-frame counted by the PAPA camera as a function of stellar magnitude B. In each case the bandwidth is fixed to 110 nm, but the aperture size and exposure time are different. The topmost line shows the response for a 4 ms frame time, 60 mm apertures, and a Detective Quantum Efficiency (DQE) four times greater than the present camera. The B magnitudes of α PsA and α Eri are included for reference.

Replacement lenses

The flat-field images of the PAPA camera would be improved if several of its array lenses were replaced. The ability to correct for image artifacts was limited by the quality of some lenses. They do not all image well over the full field defined by the image tube. The poorest lenses were allocated to the coarsest masks to reduce the number of alignment artifacts. The most noticeable errors could be corrected if the lenses for the X1, Y2, and Y3 masks were replaced.

12.2 Group Delay Tracking

Time was only available for the initial testing of the fringe tracking servo. Changes to the path compensating system, and other instrumental tests at SUSI precluded further work. The next step would be to properly tune the servo and perform tests at longer baselines.

The 80 m baseline at SUSI is the longest operational baseline of any existing stellar interferometer, and within the next year baselines up to 160 m will become available. No one knows how the atmosphere affects fringe motions at baselines longer than 31.5 m, and there is clearly an opportunity with group delay tracking for a significant contribution in this field. This work only became possible after the changes to the path compensator in the later half of 1993, and has yet to be exploited.

Appendices

Appendix A

Photographs

Photo 1. The Sydney University PAPA camera, shown here on the ‘red’ table at SUSI prior to installation. The silver box is the thermoelectric cooler which contains the image tube. The collimating lens and lens-mask assembly are contained within the A frames in the center, and the analog electronics are visible at the back. They surround the photomultiplier tubes.

Photo 2. The Analog Electronics. There are three circuit boards in these electronics: one board to condition the strobe signal, and one to latch the data in each of the x and y axes. Again, the photomultipliers are hidden from view.

Photo 3. Close-up of the Lens-Mask Assembly. This and the following photograph show different views of the lens-mask assembly. The collimating lens is on the left and the array lenses are in the center. To the right the individual masks on the mask-plate are visible.

Photo 4. Overview of the Lens-Mask Assembly. The collimating lens now appears in the center and the output of the image tube is just out of view on the left. The analog electronics are on the right.

Appendix B

Design of the Gray Coded Mask Plate

The mask plate discussed in §4.2 is described in detail. This description is identical to the one used to specify the masks for the manufacturer.

B.1 A Description of the Hexagonal/Square Array

The layout of the mask plate is shown in Fig. B.1. Each circle in the figure represents the boundary of a mask. All circles are 18mm in diameter and have their centers 18mm apart so that they are touching. It is opaque in the space between circles and also opaque in the area outside the circles. The mask boundaries, in some cases, could have been square or hexagonal, but a circular boundary was chosen because it was simpler to describe.

Within the circles there is a mask pattern which is different for each mask. There are 17 different masks etched on a single glass plate. There is one *strobe* mask, eight *x* masks, and eight *y* masks. The masks are arranged in five rows, with the central three rows containing 13 masks in a hexagonal close-packed configuration. There are two masks in the first row, and two in the fifth row, which are located symmetrically in a square grid with their neighboring rows.

B.2 Mask Descriptions

The central circle of Fig. B.1 represents the *strobe* mask: it is completely transparent within the boundary of that circle.

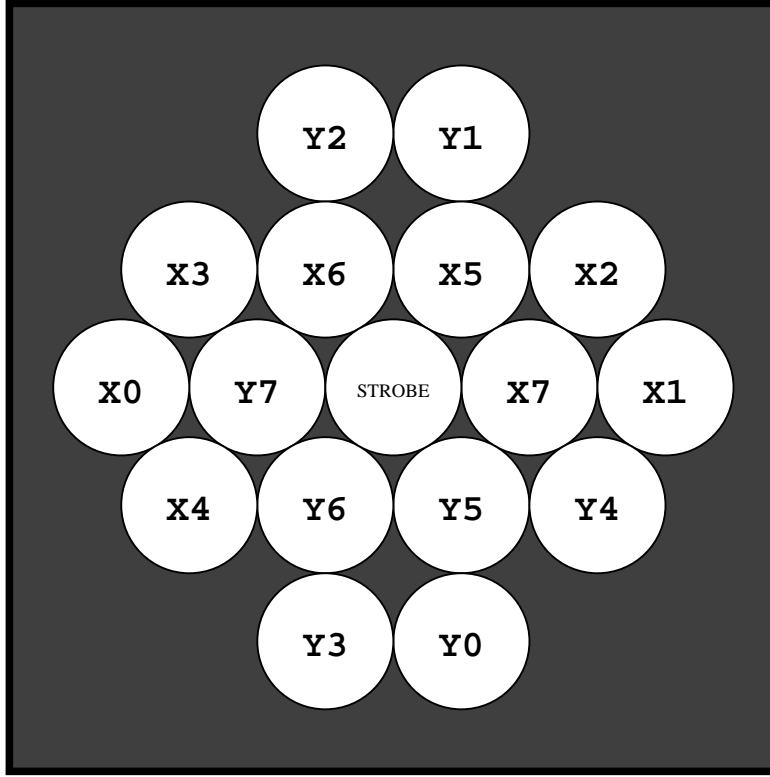


Figure B.1: The layout of the Gray coded masks. Note that the only masks that are completely surrounded, are the three central masks: X7, Y7 and the strobe.

The x masks and y masks are Gray coded to span an image that is $9984\mu\text{m}$ in diameter. This image size corresponds to the inner 20mm diameter of the intensifier, imaged with a magnification of .5 through the collimating lens and zoom-lens. It is $9984\mu\text{m}$ rather than $10000\mu\text{m}$, since 256 divides evenly into 9984, and thus the mask dimensions for each mask come out in microns rather than fractions-of-microns.

The 0 order mask divides the image in two; the 1st order mask has a central clear stripe $4992\mu\text{m}$ wide, $1/2$ of the image size; the 2nd order mask has a central opaque stripe $2496\mu\text{m}$ wide, $1/4$ of the image size; and so on until the 7th order mask which has a central opaque stripe $78\mu\text{m}$ wide, $1/128^{\text{th}}$ of the image size. The corresponding pixel size on the mask is therefore $39\mu\text{m}$, $1/256^{\text{th}}$ of the image size, and the pixel size at the image tube is $78\mu\text{m}$. The masks are described in more detail in the accompanying figures and table.

The masks are composed of alternating opaque and transparent stripes. The boundaries of the x stripes are all parallel to each other, and the boundaries of the y stripes are all parallel to each other. However, the y stripes are perpendicular to the x stripes. The width of these stripes is dependent on the *order* of the mask; masks of a higher order have smaller stripe widths—half the width of the previous order. The

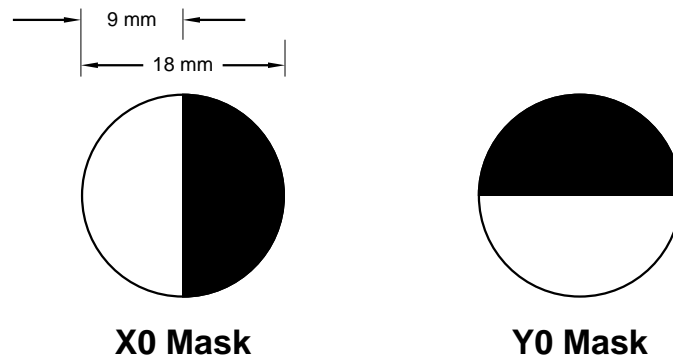


Figure B.2: **The X0 and Y0 Masks:** Half of the mask is opaque, as shown above. For the X0 mask the boundary between opaque and transparent regions is the diameter that runs vertically across the circle; it is transparent to the left of the diameter, and opaque to the right. The Y0 mask is the same as the X0 mask, except that it is rotated 90 degrees *counter-clockwise*.

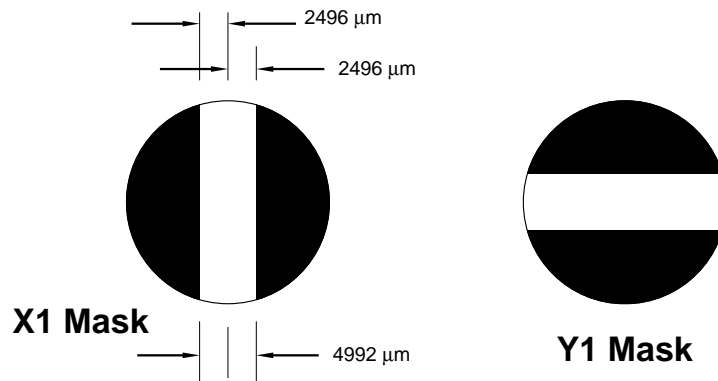


Figure B.3: **The X1 and Y1 Masks:** In the X1 mask there is one central stripe which is *transparent*. That stripe is $4992\mu\text{m}$ wide, and is centered on the diameter that runs vertically across the circle. The rest of that mask is opaque. The Y1 mask is the same as the X1 mask, except that it is rotated 90 degrees. The X1 and Y1 masks are both shown above.

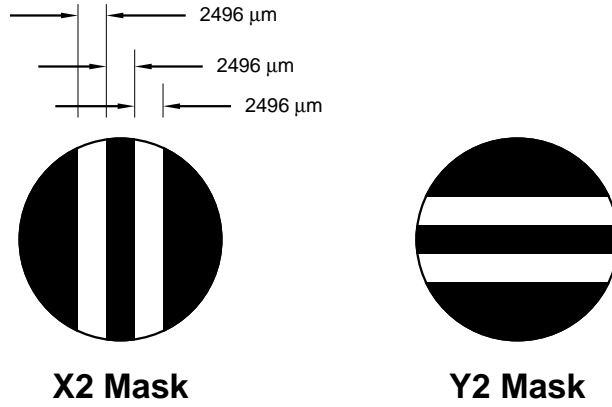


Figure B.4: **The X2 and Y2 Masks:** For the X2 mask the central stripe is *opaque* and $2496\mu\text{m}$ wide, centered on the diameter that runs vertically through the circle. On either side of that stripe there are transparent stripes, each $2496\mu\text{m}$ wide. The rest of the mask is opaque. The Y2 mask is the same as the X2 mask, except that it is rotated 90 degrees. The X2 and Y2 masks are depicted above.

x and y masks of a given order are identical, except that the y masks are rotated 90 degrees (counterclockwise) with respect to the x masks.

The masks are made *over-sized*. Although the image is only 10mm in diameter the masks themselves extend to a diameter of 18mm. Having the masks oversized makes the camera alignment easier: all masks are periodic, and so it doesn't matter what period is used as a reference in the alignment. The masks of order 0, 1, and 2 are *not* made entirely periodic because of the large period of their stripes. Masks of order 3 and higher all have patterns which are made to extend to fill the circle in which they are centered.

The width a of the stripes in each mask is given in table B.2 and depicted in Fig. B.5.

Mask #	Width of Stripes a		
X3 & Y3	$9984/8$	$=$	$1248\ \mu\text{m}$
X4 & Y4	$9984/16$	$=$	$624\ \mu\text{m}$
X5 & Y5	$9984/32$	$=$	$312\ \mu\text{m}$
X6 & Y6	$9984/64$	$=$	$156\ \mu\text{m}$
X7 & Y7	$9984/128$	$=$	$78\ \mu\text{m}$

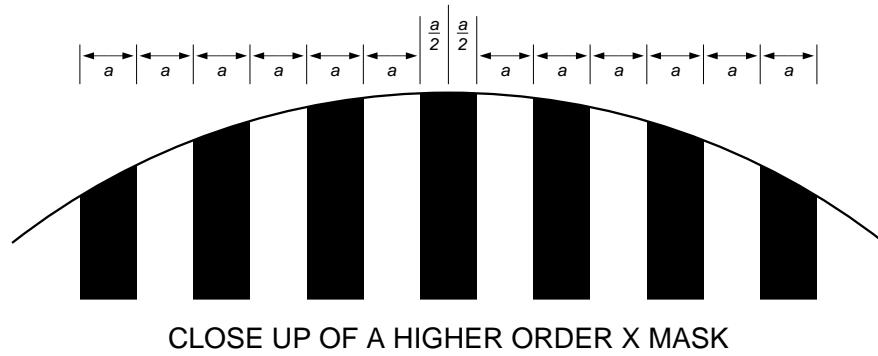


Figure B.5: **X and Y Masks of order 3, 4, 5, 6, and 7** : The x and y masks of order 3, 4, 5, 6, and 7 may be described in the following manner: each of these masks has a central stripe which is *opaque*; the central stripe is symmetric about a diameter of the circle; clear and opaque stripes alternate across the submask; clear and opaque stripes are of the same width in any one mask; the pattern that describes a mask is extended until the boundary of the circle that bounds that mask. The Y masks appear identical to the X masks except that they are rotated 90 degrees.

B.3 The Mask Plate

The full scale mask plate is shown in Fig. 4.2. The lower order masks were located to shield against vignetting, as described in §4.2. The X7 and Y7 masks were placed either side of the strobe. As can be seen in the design of the lens holder, Fig. 4.3, there is no lateral access to these lenses. However, they were used as the reference in the alignment and could be fixed in place after being focused.

The mask was manufactured by Precision Photomask Inc., 4950 Fisher St., Montreal, PQ, Canada, H4T 1J6. They defined it using their Optical Pattern Generation method, which can define structures as small as $3\mu\text{m}$ in size. This method maintains an angular tolerance $\pm 0.1^\circ$, a position accuracy of $\pm 0.6\mu\text{m}$ over a six inch square area, and has a standard dimensional tolerance of $\pm 1\mu\text{m}$ for structures larger than $25\mu\text{m}$. The mask was made using anti-reflection chrome deposited on sodalime glass. The pattern is centered on a glass plate 4 inches square, and 0.09 inches thick. They retain the master plate and will manufacture further copies upon request.

Appendix C

Alignment of the PAPA Camera

The images on each mask must all have the same magnification, and must all lie at appropriate locations if the addresses are to be decoded correctly. This was discussed at length in Chapter 3. The alignment procedure will now be presented.

C.1 Tilt Plates and Lens Rotation

There are two methods that have been used for displacing images to obtain an alignment: the rotation of tilt plates and the rotation of lenses. Tilt plates can be used to laterally shift the image and move it in a circle. However, moving or rotating a lens has the same effect without introducing image aberrations.

C.1.1 Tilt-plates

Tilt-plates were used in the design by Adaptive Optics Associates (1988). A tilt-plate is a plane parallel plate of glass which is set at an angle within an optical system. If a collimated beam passes through it then it will be shifted, or offset, in a direction normal to its direction of propagation. The amount of offset ρ is dependent upon the glass thickness d , the tilt angle θ_a , and the index of refraction of the glass n_g .

$$\rho = d \frac{\sin(\theta_a - \theta_g)}{\cos \theta_g}, \quad \text{with} \quad n_a \sin \theta_a = n_g \sin \theta_g. \quad (\text{C.1})$$

Here n_a is the index of refraction of air, θ_g is the angle of refraction in the glass, and the equation on the right is Snell's law. We may also expand the sine term and re-write the equation in the following form,

$$\rho = d \cos \theta_a (\tan \theta_a - \tan \theta_g), \quad (\text{C.2})$$

also derived in by Smith (1990, §4.8).

If a tilt-plate is placed in a converging beam, between a lens and its image plane, then it will shift the location of the image by the amount ρ . This effect has been used in rangefinders and optical micrometers.

Unfortunately, tilt-plates will introduce aberrations in the image. The glass not only shifts the focus laterally, but also displaces it further along the optical axis. The rays in the converging beam that arrive at larger angles are displaced slightly further along the optical axis, and this results in spherical aberration (Smith, 1990). To a lesser degree there also exist other effects: chromatic aberration, astigmatism, sagittal coma, and lateral chromatic aberration. Smith lists formulae to describe each of these terms, and in each case the aberrations are shown to scale linearly with plate thickness d .

In the Standley-Nisenson camera the following parameters were used: tilt angle $\theta_a = 30^\circ$, $n_g = 1.5$, and plate thickness $d = 1\text{mm}$. If we assume that $n_a = 1.0$ then we find that $\theta_g = 19.47^\circ$, and so we obtain an offset radius of $\rho = 194\mu\text{m}$. The tilt-plates were mounted at a fixed angle and were allowed to rotate through 360° . It is the rotation of the tilt-plates that provided the alignment.

If (x, y) were the coordinates of the incoming ray, then the coordinates of the outgoing ray would be (x', y') , where

$$x' = x + \rho \cos \varphi, \quad y' = y + \rho \sin \varphi, \quad (\text{C.3})$$

and φ is the angle of rotation chosen to obtain the proper alignment. Tilt-plates cause both x' and y' to change at the same time. This, however, does not pose a problem since a single mask serves only one coordinate axis; for an x mask the y offset doesn't matter, and for a y mask the x offset doesn't matter.

C.1.2 Zoom-lens rotation

The same principle may be used in a slightly different way; if we rotate the zoom-lenses then we find that the same range of adjustment is already available and that the tilt-plates are unnecessary. The centration of the zoom lenses is not perfect: the optical axes of the lenses do not necessarily match their mechanical axes, and so if the lens is rotated around its mechanical axis then the image will move in a circle in the image plane.

For the 75mm Rolyn Zoom lenses, working with a 150mm focal length collimating lens, the range of adjustment, along either the x or y axis, ranged from $\pm 80\mu\text{m}$ (which was unusually small) to $\pm 500\mu\text{m}$, with more typical values around $\pm 250\mu\text{m}$. The range is therefore typically larger than that offered by the tilt plates. It follows that if the lenses are mounted correctly then the alignment may be performed *without* the use of tilt plates. If they are omitted from the design then the image quality at

the masks will be improved, the design of the camera will be simplified, and the total cost will be slightly reduced. The Sydney University camera was aligned *without* tilt plates.

C.2 Apparatus and Preparation

The alignment technique was briefly described by Papaliolios and Mertz (1982) in their discussion of the first camera. A light source is used to illuminate selected masks in the mask-plate, and their images are projected, or rather back-projected, into the plane where the image intensifier output would be. If two masks are back-projected in this way then it is possible to view their relative alignment, and therefore to adjust them appropriately.

The apparatus consists of the following: the front end of the PAPA camera, including the collimating lens and the lens-mask assembly; a slide projector to illuminate the mask-plate; appropriate apertures at the mask-plate to select the masks that will be illuminated by the projector; a fiber-optic faceplate mounted at the location where the output of the image intensifier would be; and a travelling microscope used to examine the images on the faceplate.

C.2.1 Autocollimating the main lens

Before the alignment, and before the lens-mask assembly is installed, the collimating lens should be auto-collimated. This may be accomplished with the help of the faceplate. If the faceplate is illuminated from the front end of the camera, and a mirror placed beyond the collimating lens, then light reflected back into the lens will form an image of the surface of the faceplate. When the lens is auto-collimating, then that image will lie back on the faceplate. When the lens is focused properly it is possible, with the aid of a microscope, to peer into the faceplate and see, as a return image, the hexagonal fiber-bundles of the faceplate itself.

C.2.2 Lens-mask allocation

There may be some advantage in carefully choosing the placement of each lens. Although all of the zoom-lenses should produce high quality images, and there should be little difference between the best lens and the worst lens, in practice, however, some lenses are better than others. Therefore the best lenses should be assigned to the finest masks, the poorer lenses to the coarsest masks, and the worst lens to the strobe channel. In this way the effect of bit errors will be minimized, since errors only occur near mask boundaries, and the coarser masks have fewer mask edges.

C.3 The Alignment Procedure

A preliminary adjustment should first be made: the X0 lens should be rotated so that the back-projected image of the X0 mask is at some mean location with respect to all other mask images. Had the image of the X0 mask been at its extreme range then the alignment may be difficult, or impossible. It may take some time to determine the best position, however once that is accomplished then the alignment can begin.

The X0 and X7 masks should each be projected in turn onto the faceplate. It is then possible to adjust the lens of the X0 mask, with respect to the X7 image, until the central edge of the X0 mask exactly splits an opaque stripe of the X7 mask. These two masks are then aligned with respect to each other, and the center of the image is defined for the x axis.

The remaining x masks can be aligned using the X0 and X7 masks as references. Figure 3.4 contains an alignment chart to aid in this. There are six alignment figures—one for each mask from X6 to X1. In each case the given mask is flanked by the X0 mask to the right, and the X7 mask to the left, each in the proper alignment. The thin dark horizontal lines indicate the location of pixel boundaries. All 256 pixels are presented here. With these figures as a reference, the lenses of each mask can be rotated in turn until the proper alignment is achieved. Using the travelling microscope the periods of the X7 mask can be counted, with respect to the X0 mask, to insure that each mask is located correctly. After the x masks have been aligned then the y masks may be aligned in the same way.

It is now crucial that the image intensifier be correctly positioned. On installation its output surface must lie where the surface of the faceplate had been when the alignment was conducted. It is also important to be able to check the alignment with the image tube in place. This may be done in the following manner.

Set up the camera as it was for the alignment, but remove the faceplate and replace it with the image intensifier. Install an opaque mask over the mask plate having only two apertures: one at the X7 mask and one at the X6 mask. Illuminate the X7 mask so that a back-projected image of that mask is formed where the image tube should be. Then take a microscope and focus on the chrome surface of the X6 mask. You should see the X6 mask clearly defined, because you are looking at the mask itself, and you should also see an image of the X7 mask, returned from the surface of the image tube, if the image tube is near where it should be. The alignment and image quality can be checked in this way. If the masks were aligned previously then the alignment should be regained by moving the image tube axially. The best-focus and the correct alignment should occur at the same time.

By choosing masks other than X6 it is possible to check the alignment of the whole x axis. Similarly if the Y7 mask is used as a reference it is possible to check

the y axis. This serves as the final test of the PAPA camera optics, and what you see is what you get.

C.4 The Spectrum of the Image Tube Phosphor

The alignment should be carried out with the same spectrum of light emitted by the phosphor of the intensifier. This would highlight any problems with the lenses and yield a more accurate alignment. However, the P-47 phosphor, commonly used in PAPA cameras, has a spectral distribution that is blue-purple. This is in a region of the spectrum where the sensitivity of the eye is poor. Consequently, it would have been difficult to perform a visual alignment with blue light. For that reason the alignment of the Sydney University camera was done with white light.

Appendix D

List of Parts for the PAPA Camera

D.1 Introduction

The camera at Sydney University had originally been one of the five built by Standley and Nisenson (1989). It was dismantled, redesigned, and rebuilt for reasons described in Chapter 3. When it was rebuilt, all of the purchased components, with the exception of the mask plate, were re-used. This included the lenses, image tubes, thermoelectric cooler, and photomultipliers. All of these items had been chosen at Harvard University by Standley and Nisenson.

This appendix describes the purchased components and lists their sources in the bibliography. It also lists the blueprints for the machined parts, with drawing numbers quoted whenever possible. Some of the machined parts were reused, but many were either discarded, modified, or re-designed. I have indicated the changes that were made by placing symbols alongside the blueprint numbers. The symbols have the following meaning:

- ⊗ The part was discarded.
- The part was modified.
- ◇ The part was re-designed.

D.2 Camera Parts

D.2.1 1st Generation 25mm Varo Tube

Specifications

Part No. 510-3024-310

Distortion Corrected

Fiber optic input and output

Photocathode S20 — extended red, highest possible QE

Phosphor X3

Resolution > 56 lp/mm

Equivalent Background Illumination (EBI) $< 2.0 \cdot 10^{-11}$ lumen/cm²

No bright spots or field emission

No dark spots > 0.001 " diameter in central 22 mm dia

Regular fiber optics (Shear < 50 μ m)

All other applicable parameters to MILSPEC

Photocathode response:

550 nm > 44 mA/W (QE $> 10\%$)

800 nm > 20 mA/W (QE $> 3.1\%$)

D.2.2 25mm MCP 2nd Generation Varo Tube, Model 3603

Specifications

Part No. 510-3697-368

S20 Photocathode — not extended red

Cathode sensitivity 12% minimum at 550 nm

Phosphor P47

No film on MCP

Resolution > 25 lp/mm

Equivalent Background Illumination (EBI) $< 2.0 \cdot 10^{-11}$ lumen/cm²

No bright spots or field emission

No dark spots > 0.001 " diameter in central 22 mm dia

Regular fiber optics (Shear < 50 μ m)

All other applicable parameters to MILSPEC

Microchannel Plate (MCP) included in 3603 tube

Information only, not a part of specification

Single MCP with bias cut

Bias Angle 11°

Pore size $8.9\ \mu\text{m}$

Pore centering $11.8\ \mu\text{m}$ center to center

Open area ratio 57%

Exponential pulse-height distribution

D.2.3 Related Parts for Intensifiers**Machined Parts**

- 1 Generic Mounting Bulkhead, J. Hazen

Fiber Optic Coupler

- 1 Incom Inc., Fiber Optic Faceplate; Numerical Aperture 0.66 with EMA, 6-8 μm , 40mm diameter x 10mm thick.

Cold Box for Image Intensifiers

- 1 Products for Research, TE334RF Water Heat Exchange Thermoelectric Cooler

D.2.4 Camera Base**Machined Parts**

- 1 PD 1023 Front Bulkhead
- 4 PD 1009 Block, Strut Mount
- 4 PD 1004 Main Strut \square
- 4 PD 1005 Invar Post, Collimating Lens Mount
- 1 PD 1010 Clamp, Collimating Lens Mount
- 1 PD 1011 Ring, Collimating Lens Mount
- 1 PD 1008 Tube, Collimating Lens Mount
- 1 PD 1016 Rear Bulkhead $\otimes \diamond$

Collimating Lens

- 1 Isco - Göttingen, F 1.6/150mm - 5.91 in. Cinelux Xenon lens.

D.2.5 Lens-Mask Assembly**Machined Parts**

- 1 PD 1015 Plate, Focusing Lens & Mask Mount $\otimes \diamond$
- 1 PD 1014 Ring, Photomask Unit \otimes
- 2 PD 1019 Mount, Mask Adj. Screw \otimes
- 1 PD 1018 Ear, Mask Adj. \otimes
- 4 PD 1006 Invar Post, Focusing Lens Mount \square
- 1 PD 1007 Plate, Focusing Lens Mount $\otimes \diamond$
- 21 PD 1012 Tube, Tilt Window \otimes

Zoom Lenses

- 21 Rolyn Optics Company, 35.0075 Variable focal length lens
75–80 mm focal length, adjustable

Grey Coded Mask Plate

Advanced Reproductions, Array of 21 Gray coded masks on 110 mm
diameter AR chromed glass plates, cut to 110mm diameter. \otimes

Precision Photomask Inc., Array of 17 masks described in Appendix B.

D.2.6 Camera Back End**Machined Parts**

- 1 PD 1021 Plate, PMT Tube Mounting \square
- 1 PD 1013 Delrin Housing, PMT's $\otimes \diamond$
- 42 PD 1025 Lens & PMT Retainers \otimes
- 21 PD 1027 Field Stops, PMT Masks \otimes
- 4 PD 1017 Rear Strut \otimes
- 1 PD 1022 Rear Cover \otimes

Field Lenses

- 21 Edmund Scientific, P32,005 15mm diameter, 22.5mm focal
length, coated plano convex lenses.

Photomultiplier Tubes

- 21 Hamamatsu Corporation, R647-04 Photomultiplier Tubes.
Head-on type, 1/2 inch diameter, for photon counting.
Bialkali photocathode with peak response at 420 nm.
- 21 Hamamatsu Corporation, E849-36 Photomultiplier Sockets

D.3 List of Mechanical Drawings

Jeff Hazen designed the machined parts for the Standley and Nisenson PAPA camera. Most of the drawings listed here are his. Note that drawing PD 1002 does not exist; it was a redesign of PD 1001 that was never used. The other drawings are those of *Products for Research* which describe their Image Tube Cold Box and its rear mounting plate. The symbols have meanings as indicated at the beginning of this appendix.

Drawings by Jeff Hazen

1	PD 1001 Intensifier Potting Tube, Resdel Corp. ⊗
1	PD 1003 Flange for Potting Tube
4	PD 1004 Main Strut □
4	PD 1005 Invar Post, Collimating Lens Mount*
4	PD 1006 Invar Post, Focusing Lens Mount* □
1	PD 1007 Plate, Focusing Lens Mount ⊗ ◇
1	PD 1008 Tube, Collimating Lens Mount
4	PD 1009 Block, Strut Mount
1	PD 1010 Clamp, Collimating Lens Mount
1	PD 1011 Ring, Collimating Lens Mount
21	PD 1012 Tube, Tilt Window ⊗
1	PD 1013 Delrin Housing, PMT's ⊗ ◇
1	PD 1014 Ring, Photomask Unit ⊗
1	PD 1015 Plate, Focusing Lens & Mask Mount ⊗ ◇
1	PD 1016 Rear Bulkhead ⊗ ◇
4	PD 1017 Rear Strut ⊗
1	PD 1018 Ear, Mask Adj. ⊗
2	PD 1019 Mount, Mask Adj. Screw ⊗
21	PD 1020 Windows made of fused silica ⊗
1	PD 1021 Plate, PMT Tube Mounting □
1	PD 1022 Rear Cover ⊗
1	PD 1023 Front Bulkhead
1	PD 1024 Main Cover ⊗
42	PD 1025 Lens & PMT Retainers†⊗
1	PD 1026 Access Cover ⊗
21	PD 1027 Field Stops, PMT Masks ⊗
1	Generic Mounting Bulkhead

Products for Research Cold Box Drawings

1	TE333 + TE334 Overall Dimensions
1	55043.0040 Front Plate
1	55043.0050 Rear Plate
1	55043.0080 Rear Mounting Plate, Cooler
1	55043.0090 Rear Sleeve

† The lens and PMT retainers, PD 1025, were *not* machined. They were cut from stock material.

* The curves indicated on these drawings are not functional, and were omitted in the cameras made by CfA.

D.4 Notes and Comments**D.4.1 The Image Intensifiers**

It had been intended to use both image tubes together, in the manner suggested by Papaliolios et al. (1985) and Latham (1982). However, numerous engineering problems precluded their use, and ultimately only a single Gen II tube was used.

D.4.2 The Collimating Lens

The collimating lenses were bought through Schneider Corporation of America, and were made in Germany by Isco. They were designed for the projection of drive-in movies: **Cinelux** means cinema-light. Understandably, that market has dried up, and the lenses are no longer manufactured.

Sources of PAPA Camera Parts

VARO, Inc., Electron Devices Division, 2203 W. Walnut Street, P.O. Box 469014, Garland, Texas 75046-9014. Tel: (214) 487-4100, Fax: (214) 487-4265, Telex: 163165 VAROI UT.

Incom Inc., 205 Chapin Street, Southbridge, MA 01550-0528. Tel: (617) 765-9151.

Products for Research, 88 Holten Street, Danvers, Massachusetts 01923. Tel: (617) 774-3250, Fax: (617) 245-1628, Telex: 94-0287 (PHOTOCOOL DARS).

Schneider Corporation, 400 Crossways Park Drive, Woodbury, New York 11797. Tel: (516) 496-8500, Fax: (516) 496-8524, Telex: 960102.

Rolyn Optics Company, 706 Arrowgrand Circle, Covina, California 91722-2199. Tel: (818) 915-5707, Telex: 67-0380.

Hamamatsu Corporation, 360 Foothill Road, Box 6910, Bridgewater, NJ 08807-0910. Tel: (201) 231-0960, Fax: (201) 231-1539, Telex: 833-403.

Advanced Reproductions, 100 Flagship Drive, North Andover, MA 01845. Tel: (617) 685-2911.

Precision Photomask Inc., 4950 Fisher St., Montreal, P.Q., Canada H4T 1J6. Tel: (514) 737-7030. Fax: (514) 737-9893.

Edmund Scientific, 101 E. Gloucester Pike, Barrington, NJ 08007. Tel: (609) 547-3488. Fax: (609) 573-6295. Telex: 831-564.

Optical Radiation Corporation, 1300 Optical Drive, Azusa, California 91702. Tel: (818) 969-3344, Telex: TWX 910-584-4851.

Appendix E

The Prism Spectrometer

E.1 Prism Equations

The angle of deviation γ of rays leaving a prism is given by the relation

$$\gamma = \theta_i + \arcsin \left[(\sin \alpha) (n^2 - \sin^2 \theta_i)^{1/2} - \sin \theta_i \cos \alpha \right] - \alpha \quad (\text{E.1})$$

where θ_i is the angle of incidence of the light measured with respect to the normal of the prism surface, α is the prism angle, and the index of refraction n is a function of wavelength (Hecht, 1987). The rate of change of the angle of deviation as a function of the index of refraction is therefore

$$\frac{\delta \gamma}{\delta n} = \left[1 - \left\{ (\sin \alpha) (n^2 - \sin^2 \theta_i)^{1/2} - \sin \theta_i \cos \alpha \right\}^2 \right]^{-1/2} \frac{n \sin \alpha}{(n^2 - \sin^2 \theta_i)^{1/2}}. \quad (\text{E.2})$$

If the spectrum is mapped onto an image plane whose coordinate in the direction of the dispersion is x , then by using a thin lens of focal length f the mapping will be

$$\frac{\delta x}{\delta \lambda} = f \frac{\delta \gamma}{\delta \lambda}, \quad \text{where} \quad \frac{\delta \gamma}{\delta \lambda} = \frac{\delta \gamma}{\delta n} \cdot \frac{\delta n}{\delta \lambda}. \quad (\text{E.3})$$

We could also express the mapping in terms of the wavenumber $\kappa = 1/\lambda$ to obtain

$$\frac{\delta x}{\delta \kappa} = f \frac{\delta \gamma}{\delta n} \cdot \frac{\delta n}{\delta \lambda} \cdot \frac{\delta \lambda}{\delta \kappa}, \quad \text{with} \quad \frac{\delta \lambda}{\delta \kappa} = -\frac{1}{\kappa^2}. \quad (\text{E.4})$$

If we operate the prism at the angle of minimum deviation then the angle of incidence θ_i is chosen so that

$$\sin \theta_i = n \sin(\alpha/2), \quad (\text{E.5})$$

where n is calculated at the center of the spectral band.

E.2 Calibration of the Spectrometer

The spectrometer may be calibrated using the lines from a low pressure mercury vapour lamp. If the spectrometer has a lens of focal length f , and x is a coordinate in the image plane, and ρ is the number of pixels per mm across the detector, then we may write the mapping from wavelength to pixel as follows:

$$p(\lambda) = p_0 + \rho f \arcsin \left[(\sin \alpha) \left(n^2 - \sin^2 \theta_i \right)^{1/2} - \sin \theta_i \cos \alpha \right], \quad (\text{E.6})$$

where p_0 is an offset to be determined in the calibration.

The prism used in the spectrometer was made of SF52 glass and has an index of refraction given by a fifth order polynomial, here quoted from the Schott Glass catalog.

$$n^2 = A_0 + A_1 \lambda^2 + A_2 \lambda^{-2} + A_3 \lambda^{-4} + A_4 \lambda^{-6} + A_5 \lambda^{-8}, \quad (\text{E.7})$$

where λ is in units of μm and the coefficients for SF52 glass are $A_0 = 2.7576175$, $A_1 = -9.7069196 \cdot 10^{-3}$, $A_2 = 2.9008907 \cdot 10^{-2}$, $A_3 = 1.6834427 \cdot 10^{-3}$, $A_4 = -8.6832245 \cdot 10^{-5}$, and $A_5 = 1.1266792 \cdot 10^{-5}$. This equation is good for wavelengths between 400 nm and 750 nm to an accuracy of $\pm 3 \cdot 10^{-6}$.

The prism is first set up for operation at the angle of minimum deviation. This is accomplished by slowly rotating the prism while observing the location of a line in the spectrum imaged in the focal plane. The angle θ_i is then known. The scaling factor ρf can also be found experimentally by observing two spectral lines, and calculating their separation in pixels.

$$\begin{aligned} \frac{p(\lambda_2) - p(\lambda_1)}{\rho f} &= \arcsin \left[(\sin \alpha) \left(n_{\lambda_2}^2 - \sin^2 \theta_i \right)^{1/2} - \sin \theta_i \cos \alpha \right] \\ &\quad - \arcsin \left[(\sin \alpha) \left(n_{\lambda_1}^2 - \sin^2 \theta_i \right)^{1/2} - \sin \theta_i \cos \alpha \right]. \end{aligned} \quad (\text{E.8})$$

As a final step the offset p_0 can be determined from the location of a single spectral line:

$$p_0 = p(\lambda) - \rho f \arcsin \left[(\sin \alpha) \left(n^2 - \sin^2 \theta_i \right)^{1/2} - \sin \theta_i \cos \alpha \right]. \quad (\text{E.9})$$

From our previous equation it is possible to determine the frequency at any pixel. The only quantity that is wavelength dependent is n the index of refraction. n can be written as follows:

$$n = \left[\left\{ \frac{1}{\sin \alpha} \sin \left(\frac{p_n - p_0}{\rho f} \right) + \frac{\sin \theta_i \cos \alpha}{\sin \alpha} \right\}^2 + \sin^2 \theta_i \right]^{1/2} \quad (\text{E.10})$$

and then we may find the wavelength by placing our value of n in the equation for the index of refraction and solving for λ . For Cauchy's dispersion formula this would have been straightforward, but for the above polynomial it involves a zero search using Newton's Method (Press et al., 1992).

Appendix F

Spectrum Analysis and the DFT

The Discrete Fourier Transform (DFT) is an approximation to the continuous Fourier transform. It operates with a number of implied assumptions that are important to understand. It is the purpose of this section to give a brief overview of Fourier transform theory and to provide some insight into the use of the Discrete Fourier Transform. The Fast Fourier Transform (FFT) is an efficient mathematical technique of evaluating the DFT. It will not be reviewed here. For more information on the history and algorithm of the FFT the reader is referred to the volume of collected papers edited by Rabiner and Rader (1972).

The Fourier transform is discussed in detail in numerous texts and papers; see for instance Bracewell's *The Fourier Transform and its Applications* (Bracewell, 1986). In the following section Bracewell's notation, particularly the $\Pi(x)$ and $\text{III}(x)$ functions, will be used, but the form of the Fourier transform will be in keeping with the discussions contained in Bergland (1969), and Kay and Marple (1981). Here we will introduce the transform as a basis of discussing the DFT.

F.1 The Fourier Transform

The time varying function $x(t)$ can be expressed as a sum, or integral, of sine and cosine waves, $\exp(j2\pi ft)$, of different frequencies f , each with different amplitudes $|X(f)|$, and phase lags $\arg X(f)$. The function $X(f)$ is complex valued, and is said to be the frequency domain representation of $x(t)$. Both $x(t)$ and $X(f)$ should be viewed as alternate representations of the same function. The Fourier transform relationship, relating these two, is as follows:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \quad \text{and} \quad x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df$$

$X(f)$ can then be understood as the amplitude and phase derived from the projection of $x(t)$ onto a wave $\exp(-j2\pi ft)$ of frequency f . It is important to note that $x(t)$ and $X(f)$ are continuous functions which may be of infinite extent in their respective domains.

The squared modulus of the Fourier transform $|X(f)|^2$ is often termed the spectrum of $x(t)$. This represents an Energy Spectral Density (ESD), and if it is time-averaged it becomes the Power Spectral Density (PSD). As well, the Wiener-Khinchin theorem, applied to a function $x(t)$, relates its autocorrelation $R_{xx}(\tau)$ to its PSD (Kay and Marple, 1981).

F.2 Some Useful Functions and Fourier Transforms

The Discrete Fourier Transform can be modelled in terms of a continuous Fourier transform by using appropriate functions and relationships, which will now be reviewed.

F.2.1 Fourier transform relationships

If $h(t)$ and $H(f)$ are Fourier transform pairs, then we may express that relationship as follows:

$$h(t) \rightleftharpoons H(f)$$

If $g(t)$ and $G(f)$ are also transform pairs, $g(t) \rightleftharpoons G(f)$, then the following relationships are true:

$$\begin{aligned} h(t/a) &\rightleftharpoons |a| H(fa) \\ h(t-a) &\rightleftharpoons e^{-j2\pi af} H(f) \\ h(t) + g(t) &\rightleftharpoons H(f) + G(f) \\ h(t) * g(t) &\rightleftharpoons H(f) G(f) \\ h(t) g(t) &\rightleftharpoons H(f) * G(f) \end{aligned}$$

where the asterisk ‘*’ is used to denote a convolution.

$$h(t) * g(t) \equiv \int_{-\infty}^{\infty} h(\tau) g(t - \tau) d\tau. \quad (\text{F.1})$$

F.2.2 The Rectangle function, $\Pi(t)$

The rectangle function $\Pi(t)$ is zero everywhere except for a period of unit length near the origin, where it is of unit height.

$$\Pi(t) = \begin{cases} 0, & |t| > 1/2 \\ 1, & |t| < 1/2 \end{cases}$$

It is most often seen in a product with other functions, so that those functions will then have non-zero values only over a given period.

$$\Pi\left(\frac{t}{\Delta t}\right) h(t) = \begin{cases} 0, & |t| > \Delta t/2 \\ h(t), & |t| < \Delta t/2 \end{cases}$$

The Fourier transform of the rectangle function is the sinc function.

$$\Pi(t) \Leftrightarrow \text{sinc}(f), \quad \text{where} \quad \text{sinc}(f) = \frac{\sin \pi f}{\pi f} \quad (\text{F.2})$$

Therefore a function $h(t)$ that is truncated by the rectangle function will have its spectrum smoothed, or convolved, by a sinc function.

$$\Pi\left(\frac{t}{\Delta t}\right) h(t) \Leftrightarrow |\Delta t| \frac{\sin \pi f \Delta t}{\pi f \Delta t} * H(f)$$

F.2.3 The Sampling or Replicating function, $\text{III}(t)$

The sampling function $\text{III}(t)$, or *Sha* function, is an infinitely long series of delta functions, where each delta function has the property

$$\delta(t - n) = \begin{cases} 1, & t = n \\ 0, & \text{otherwise} \end{cases}$$

and is located at regular intervals in t .

$$\begin{aligned} \text{III}(t) &= \sum_{n=-\infty}^{\infty} \delta(t - n) \\ \frac{1}{\Delta t} \text{III}\left(\frac{t}{\Delta t}\right) &= \sum_{n=-\infty}^{\infty} \delta(t - n\Delta t) \end{aligned}$$

Any function that is multiplied by the $\text{III}(t)$ function becomes sampled, so we may write

$$\frac{1}{\Delta t} \text{III}\left(\frac{t}{\Delta t}\right) h(t) = \begin{cases} h(t), & t = n\Delta t \\ 0, & \text{otherwise} \end{cases}$$

The sampling function is its own Fourier transform,

$$\text{III}(t) \Leftrightarrow \text{III}(f).$$

Therefore, any function that is multiplied by the $\text{III}(t)$ function, that is to say sampled, has a spectrum which is replicated at regular intervals in frequency space; a spectrum that is periodic.

$$\frac{1}{\Delta t} \text{III}\left(\frac{t}{\Delta t}\right) h(t) \Leftrightarrow \text{III}(f\Delta t) * H(f)$$

F.3 A Note on Sampling Theory

If the function $h(t)$ is bandlimited to a frequency of f_w then we can write

$$h(t) \Leftrightarrow H(f)$$

$$H(f) = \Pi\left(\frac{f}{2f_w}\right) H(f)$$

where $H(f)$ has positive and negative frequency components as high as f_w . We may completely reconstructed $h(t)$ from samples taken at intervals of

$$\Delta t = \frac{1}{2f_w},$$

where the frequency $2f_w$ is known as the Nyquist sampling frequency, and the reconstruction is described by Shannon's sampling theorem. A tutorial on sampling theory is given by Jerri (1977); the following is an illustration using Fourier transform theory.

F.3.1 The reconstruction of a sampled signal

The sampling theorem may be illustrated as follows: A function $h(t)$ that has its spectrum truncated by the rectangle function will be smoothed, or convolved, by a sinc function.

$$2f_w \frac{\sin 2\pi t f_w}{2\pi t f_w} * h(t) \Leftrightarrow \Pi\left(\frac{f}{2f_w}\right) H(f).$$

We know that sampling in the time domain will cause a periodic replication of the spectrum in the frequency domain. If the sampling frequency is at least $2f_w$ then the replicated spectra will be non-overlapping, and that part of the resultant spectrum which lies in the interval of $-f_w < 0 < f_w$ will represent the true spectrum of the (unsampled) signal. The sampling and its effects in the frequency domain may be written as

$$\frac{1}{\Delta t} \text{III}\left(\frac{t}{\Delta t}\right) \left[2f_w \frac{\sin 2\pi t f_w}{2\pi t f_w} * h(t) \right] \Leftrightarrow \text{III}\left(\frac{f}{2f_w}\right) * \left[\Pi\left(\frac{f}{2f_w}\right) H(f) \right]$$

where $\Delta t = 1/(2f_w)$. The sampled function fully reproduces the bandlimited spectrum, since we have

$$\text{III}\left(\frac{f}{2f_w}\right) * \left[\Pi\left(\frac{f}{2f_w}\right) H(f) \right] = H(f), \quad -f_w \leq f \leq f_w.$$

So, although sampling makes the spectrum periodic it is still possible to reproduce the original signal by putting the data through a low pass filter whose cut-off frequency is f_w .

F.4 The Discrete Fourier Transform

The Discrete Fourier Transform is an approximation to the integral Fourier transform. The DFT requires that both the time domain and frequency domain functions be represented by a series of samples that are of finite duration. We therefore replace $x(t)$ by the series of samples x_n where n is an index which numbers the samples from 0 to some large number $N-1$; the samples being taken at regular intervals Δt apart. We also replace $X(f)$ by the series of samples X_m , where m is an index whose maximum value is the integer $N-1$, and the samples lie at intervals of $\Delta f = 1/(N\Delta t)$. We can now write the variables t and f as

$$t = n\Delta t, \quad n = 0, 1, 2, \dots, N-1 \quad \text{and} \quad f = \frac{m}{N\Delta t}, \quad m = 0, 1, 2, \dots, N-1$$

and the product ft , which appears in the Fourier transform integral, becomes

$$ft = \frac{nm}{N}.$$

The Discrete Fourier Transform is written as follows:

$$\begin{aligned} X_m &= \Delta t \sum_{n=0}^{N-1} x_n e^{-j2\pi nm/N} \\ x_n &= \frac{1}{N\Delta t} \sum_{m=0}^{N-1} X_m e^{j2\pi nm/N} \end{aligned}$$

In most representations of the DFT the term Δt is dropped, but it has been included here for clarity. The DFT analog of the Energy Spectral Density is known as the *Periodogram*.

$$\mathcal{P}_m = \frac{1}{N\Delta t} |X_m|^2$$

The spectrum represented by $|X_m|^2$ is no longer the true spectrum $|X(f)|^2$, but rather an approximation. It is important to understand how these two functions differ.

F.5 Artifacts Produced by the DFT

Let us now represent the differences between the integral transform and the discrete Fourier transform using the rectangle and sampling functions discussed earlier. It must be remembered that the function $x(t)$ was truncated and sampled, to produce x_n , before it was transformed. The truncation means that the time series has an abrupt beginning and an abrupt end, and because we have sampled $x(t)$ the space between the samples is assumed to be zero valued. We have changed $x(t)$ in the following way:

$$x(t) \longrightarrow \underbrace{\left[\Pi\left(\frac{t}{N\Delta t}\right) \right]}_{\text{windowed}} \underbrace{\left[\frac{1}{\Delta t} \text{III}\left(\frac{t}{\Delta t}\right) \right]}_{\text{sampling}} x(t) \underbrace{* \text{III}\left(\frac{t}{N\Delta t}\right)}_{\text{replicated}},$$

where the quantity within the square brackets is the truncated and sampled version of $x(t)$. The *Sha* function outside of the brackets insures that the transform of this is sampled at the correct intervals in the frequency domain; the DFT models the data as if it were composed of frequency components that are harmonically related, where the highest harmonic is the $(N/2)$ th. This implies that the new version of $x(t)$ is periodic in time, with a period of $N\Delta t$. The frequency domain equivalent of $x(t)$ therefore becomes

$$X(f) \longrightarrow \left[\underbrace{N\Delta t \frac{\sin(\pi f N\Delta t)}{\pi f N\Delta t}}_{\text{smoothed}} * \underbrace{\text{III}(f\Delta t)}_{\text{replicated}} * X(f) \right] \underbrace{\frac{1}{N\Delta t} \text{III}\left(\frac{f}{N\Delta t}\right)}_{\text{sampled}}$$

The Fourier transform represented here is understandably different from $X(f)$. All of these alterations produce artifacts which appear in the spectrum, and give rise to effects known as aliasing, spectral leakage, and the picket-fence effect. They require some explanation, but let us first comment on resolution and bandwidth.

F.5.1 Resolution and bandwidth

Resolution in a spectrum is a measure of the ability to distinguish between two neighbouring spectral lines of equal intensity. In a DFT the resolution is given by the reciprocal of the extent of the samples in the time domain, $\Delta f_w = 1/(N\Delta t)$. The sampling rate determines the bandwidth of the spectrum. If the data was sampled at intervals of Δt , then the bandwidth in the spectrum, the highest frequency component that should be present, will be $f = 1/(2\Delta t)$.

F.5.2 Aliasing

The DFT produces the frequency components X_m up to a value of $m = N/2$ corresponding to half the sampling frequency, and it is assumed that no higher frequency components exist. If they do then they will be seen, after detection, to oscillate at a lower frequency, and will be summed with the true frequency components that already exist there. This effect is known as *aliasing*. It is therefore imperative that the signal $x(t)$ be filtered so that this does not happen; otherwise power will appear folded back to lower frequencies, confusing the interpretation of the spectrum.

F.5.3 Spectral leakage

The DFT uses a set of orthogonal basis functions to model an arbitrary time varying function $x(t)$. These basis functions are waves of the following form: $\exp(j2\pi nm/N)$, where m is an integer, m/N represents the wave's frequency, and n is an integer that increases along the time axis. These frequencies are therefore harmonically related.

The projection of $x(t)$ onto these functions yields the coefficients X_m for each value of m . If $x(t)$ happens to be a wave of a single frequency that matches the frequency of a basis function, then the projection will fall on one function only; all other functions will have coefficients that are identically zero. If, on the other hand, $x(t)$ is a wave of a single frequency which does *not* match any of the basis functions, but has a frequency that falls between harmonics, then its projection will yield non-zero coefficients for every function. This is known as *spectral leakage*.

It can be understood in another way. The DFT forces the time series to be periodic, and the beginning and end of the series will meet each time the series is repeated. This wrap-around discontinuity gives rise to frequency components that would not exist in the integral transform $X(f)$. It means that if the sampled function is not periodic over the length $N\Delta x$, implying that there exist frequency components that are not sampled, then the power in those components will be divided amongst the nearest neighbouring harmonics.

Spectral leakage is an artifact of the truncation and windowing. Truncation is equivalent to a multiplication by a rectangle function $\Pi(x)$. A time series that is multiplied by $\Pi(x)$ has a spectrum that is convolved by $\text{sinc}(f)$. If the spectrum is sampled then this convolution may spread frequency information across neighbouring samples. The amount of leakage depends of the sidelobe structure of the convolving function. If the data sets are tapered, or windowed, towards the beginning and end of the series, then it is possible to reduce the sidelobe levels, but *always* at the expense of resolution. Windowing effectively reduces the length of the data set, throwing away information, and the resolution is therefore reduced. An extensive review of the use of window functions is given by Harris (1978).

F.5.4 The picket-fence effect

Let us assume that $x(t)$ is a single sinusoid, whose frequency we can change to any frequency we wish, in a continuous fashion. Spectral leakage causes the power of that sinusoid to be distributed amongst neighbouring harmonics (values of m corresponding to X_m) when its frequency lies between harmonics. When the frequency of $x(t)$ is increased through several harmonics its peak height will appear to increase and decrease. It peaks at the harmonics where there is no spectral leakage, and reaches a minimum value half-way between them, where the leakage is the greatest. This ripple in the spectrum has been called *the picket-fence effect*.

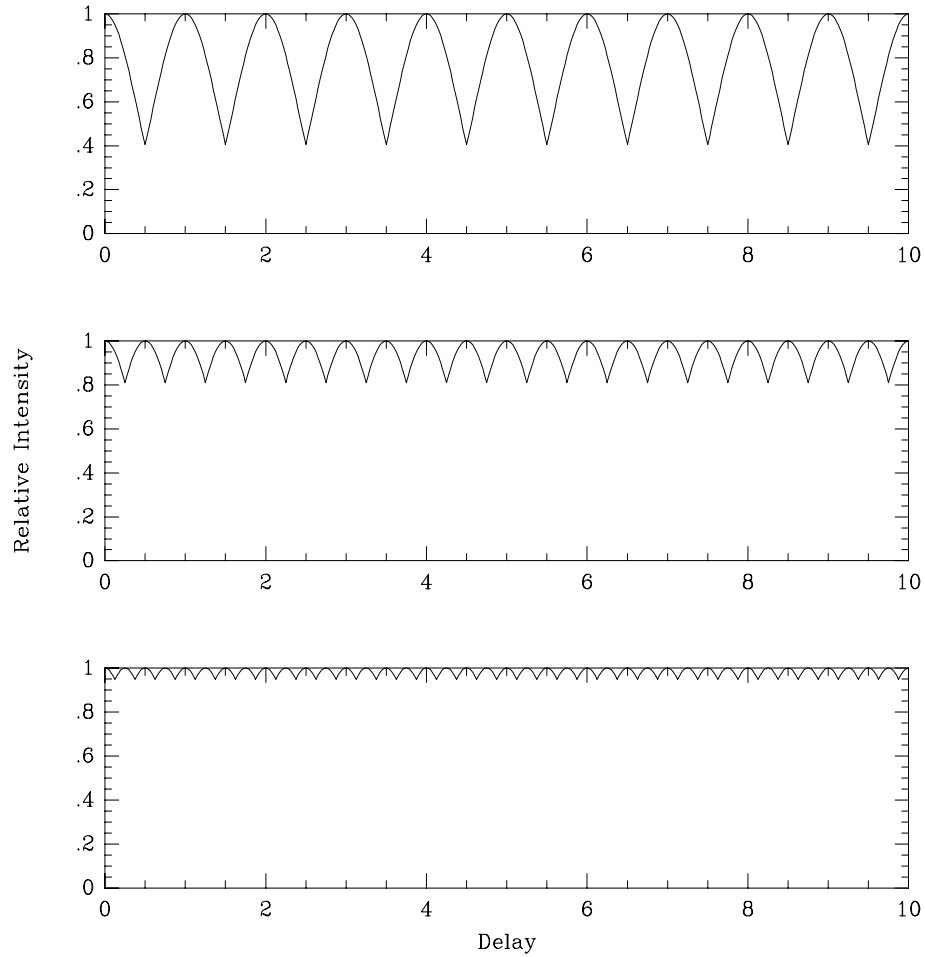


Figure F.1: The picket-fence effect: These illustrations show how the picket-fence effect alters the power-spectrum. The spectrum is sampled at the locations X_m where the curves peak. In the top diagram there is no zero-padding and the spectrum is sampled in intervals of $\Delta s = 1$. At frequencies between samples the power appears to drop off rapidly. When the length of the data set is doubled by zero-padding, then the response is that shown in the middle figure. The ‘filter’ at each sample retains the same shape, but the samples move closer together. In the bottom figure the data set has been extended to four times its original length through zero-padding, and the power drop is negligible.

If the time series $x(t)$ is artificially extended with samples that are zero valued—a procedure called *zero-padding*—then the DFT will interpolate the frequency components that lie between the original harmonics. This will reduce the picket-fence effect, but will *not* increase the resolution of the spectrum; the new values simply fill-in further values of an already smoothed function. The only way to increase the resolution is to increase the extent of samples of $x(t)$. The picket-fence effect, and the effect of zero-padding are illustrated in Fig. F.1.

Appendix G

Wave Propagation in Random Media

The theory of wave propagation through weak turbulence will be reviewed in this appendix, using the spectral representation of random variables. The theory presented here is largely based on the work of Tatarski (1961, Chapt. 7) and Ishimaru (1978, Chapt. 17). The goal of these derivations is to determine the temporal power spectra of phase fluctuations observed at a point. In Chapt. 8 this is used to describe the phase-difference power spectrum, and thereby to model the behavior of the white-light fringe position in Michelson stellar interferometry. This approach has been favored by several authors (Colavita et al., 1987; Buscher et al., 1992) because it provides a convenient framework in which to compare experiments and theory.

G.1 Introduction

Since the early 60s there have been two main approaches to the study of the propagation of electromagnetic waves through turbulent media: the classical technique of weak perturbations, and the modern methods that treat strong perturbations and multiple scattering.

Lawrence and Strohbehn (1970) state that in the 1960s most of the literature followed the Russians' use of the so-called Rytov approximation, a weak perturbation method that uses the scalar wave equation. A derivation using this technique was presented by Tatarski (1961) wherein the logarithm of the field is taken before the perturbation is applied. Fante (1975) points out that the Rytov approximation is equivalent to an integration over a series of 'phase screens' and does not account for multiple scatter from turbulent eddies. Therefore, over long horizontal paths, greater than 1 km, it does not adequately describe the propagation of electromagnetic waves.

Disagreements between the theory and observations were noted by the Russians in the mid 1960s.

The failure of the classical approach gave incentive to the development of modern techniques which attempt to overcome the limitations of the Rytov method. Although there does not exist any rigorously derived theory that explains all experimental data, the range of validity of predictions has been dramatically improved: whereas the Rytov method is limited to distances of less than about 1 km, the Markov approximation is valid for distances of hundreds or thousands of kilometers in the Earth's atmosphere (Lawrence and Strohbehn, 1970). Tatarskii (1971, Chapt. 5), and Tatarskii and Zavorotnyi (1980) present this method, which is related to quantum field theory. Further developments are summarised in the review article by Strohbehn (1978).

Despite the improvements of modern techniques, studies of atmospheric turbulence for astronomy have invariably used the classical approach. Coulman (1985) followed Tatarski's use of the Rytov approximation in a discussion of astronomical seeing. Roddier (1981) used the phase-screen method of Lee and Harp (1969) to review the theory and effects of atmospheric turbulence on astronomical observations. He argued that the saturation effects in scintillation, which mark the failure of the weak perturbation method, are only evident for observations at 60 degrees or more away from the zenith, and that experiments have suggested the phase and angle-of-arrival statistics have a greater range of validity than the log-amplitude predictions (Clifford, 1978, p.37). He concluded that although modern techniques are required to treat horizontal propagation of waves near ground under varying conditions, classical methods are still suitable for vertical propagation at good astronomical sites.

G.2 Spatial Covariances of Log-Amplitude and Phase

In Tatarski's derivation (1961, Chapt. 7) he begins by assuming that the atmosphere will cause small changes to the wave, so that the propagating wave at a point \mathbf{r} is composed of an unperturbed wave $U_0(\mathbf{r})$ and a small (complex) phase perturbation $\psi_1(\mathbf{r})$. We have therefore

$$U(\mathbf{r}) = U_0(\mathbf{r}) \exp(\psi_1(\mathbf{r})),$$

where the exponent is separated into real and imaginary parts:

$$\psi_1(\mathbf{r}) = \chi(\mathbf{r}) + jS_1(\mathbf{r}).$$

$\chi(\mathbf{r})$ represents changes in the logarithm of the amplitude, and $S_1(\mathbf{r})$ represents changes in the phase. The plane wave travels in the positive direction along the h axis, entering the medium at $h = 0$. An expression is then derived for the phase of the field at $h = L$, as an integration over the changing index of refraction.

The log-amplitude and phase have covariances

$$B_\chi(\mathbf{r}_1, \mathbf{r}_2) = \langle \chi(\mathbf{r}_1) \chi(\mathbf{r}_2) \rangle \quad \text{and} \quad B_s(\mathbf{r}_1, \mathbf{r}_2) = \langle S_1(\mathbf{r}_1) S_1(\mathbf{r}_2) \rangle, \quad (\text{G.1})$$

where \mathbf{r}_1 and \mathbf{r}_2 are the locations where the field is sampled. A field is said to be *homogeneous* if its mean value is constant and if the statistical properties of the field are independent of where they are measured. We can then shift all vectors by an amount \mathbf{r}_2 to obtain

$$B(\mathbf{r}_1, \mathbf{r}_2) = B(\mathbf{r}_1 - \mathbf{r}_2, 0).$$

Furthermore, a field is said to be *isotropic* if its properties are independent of direction, $\rho = |\mathbf{r}_1 - \mathbf{r}_2|$. In the following derivations it will be assumed that the medium is locally isotropic, but slowly changing along the h axis.

Now, if $\kappa = 2\pi/l$ is a coordinate in spatial wavenumber (proportional to spatial frequency) and $k = 2\pi/\lambda$ is the wavenumber of the electric field then the covariances can be expressed in terms of the power spectrum of the index of refraction fluctuations, $\Phi_n(h, \kappa)$:

$$B_\chi(L, \rho) = 4\pi^2 k^2 \int_0^L dh \int_0^\infty \kappa d\kappa J_0(\kappa\rho) \sin^2 \left[\frac{L-h}{2k} \kappa^2 \right] \Phi_n(h, \kappa),$$

$$B_s(L, \rho) = 4\pi^2 k^2 \int_0^L dh \int_0^\infty \kappa d\kappa J_0(\kappa\rho) \cos^2 \left[\frac{L-h}{2k} \kappa^2 \right] \Phi_n(h, \kappa),$$

where $J_0(\kappa\rho)$ is a Bessel function of the first kind of order 0. This general expression is also derived in a more physical fashion by Lee and Harp (1969, eqs. 13, 14), and is discussed at some length by Mandics, Lee, and Waterman (1973).

If we were to make the simplifying assumption that the wave propagates in a homogeneous medium, so that the power spectrum $\Phi_n(h, \kappa)$ is independent of position h , then

$$\Phi_n(h, \kappa) = \Phi_n(\kappa)$$

and the integration over h yields

$$B_{\chi,s}(L, \rho) = 2\pi^2 k^2 L \int_0^\infty J_0(\kappa\rho) f_{\chi,s}(\kappa) \Phi_n(\kappa) \kappa d\kappa. \quad (\text{G.2})$$

where

$$f_\chi(\kappa) = 1 - \frac{\sin(\kappa^2 L/k)}{\kappa^2 L/k} \quad \text{and} \quad f_s(\kappa) = 1 + \frac{\sin(\kappa^2 L/k)}{\kappa^2 L/k}$$

The $f(\kappa)$ terms have been called filter functions, because they filter the spectrum $\Phi_n(\kappa)$ to arrive at the covariances.

G.3 Temporal Power Spectra of Log-Amplitude and Phase

The temporal power spectrum of phase fluctuations is found using Taylor's hypothesis, that assumes the spatial fluctuations are 'frozen' and swept past the observer by a wind of a given velocity. In the case of a wind with a constant velocity we simply replace ρ by $v_\perp \tau$, where v_\perp is component of the wind speed that is perpendicular to the path of the propagating wave. We then obtain the covariances

$$B_\chi(L, \tau) = 4\pi^2 k^2 \int_0^\infty \kappa d\kappa J_0(\kappa v_\perp \tau) \int_0^L dh \sin^2 \left[\frac{L-h}{2k} \kappa^2 \right] \Phi_n(h, \kappa),$$

$$B_s(L, \tau) = 4\pi^2 k^2 \int_0^\infty \kappa d\kappa J_0(\kappa v_\perp \tau) \int_0^L dh \cos^2 \left[\frac{L-h}{2k} \kappa^2 \right] \Phi_n(h, \kappa).$$

The temporal frequency spectrum can then be written as the Fourier transform of the correlation functions.

$$W(\omega) = 2 \int_{-\infty}^{\infty} B(L, \tau) \cos(\omega \tau) d\tau = 4 \int_0^{\infty} B(L, \tau) \cos(\omega \tau) d\tau$$

This transform contains an integral over the Bessel function, the only time-dependent part of the correlation function:

$$\int_0^{\infty} \cos(\omega \tau) J_0(\kappa v_\perp \tau) d\tau = \begin{cases} [(\kappa v_\perp)^2 - \omega^2]^{-1/2} & \kappa v_\perp > \omega \\ 0 & \text{otherwise} \end{cases}$$

The power spectrum of phase fluctuations can therefore be written

$$W_s(\omega) = 16\pi^2 k^2 \int_{\omega/v_\perp}^{\infty} \kappa d\kappa \int_0^L dh \cos^2 \left[\frac{L-h}{2k} \kappa^2 \right] \Phi_n(h, \kappa) [(\kappa v_\perp)^2 - \omega^2]^{1/2}.$$

G.3.1 Temporal power spectra in a homogeneous medium

If the medium is homogeneous then the integration over h is straightforward, and after a substitution of variables Tatarskii (1971, §52, Eq. 33) writes

$$W(\omega) = \frac{8\pi^2 k^2 L}{v_\perp} \int_0^\infty f(\kappa) \Phi_n(\kappa) d\kappa', \quad (\text{G.3})$$

where $f_\chi(\kappa)$ and $f_\rho(\kappa)$ are as given before, and

$$\kappa = \sqrt{\kappa'^2 + \left(\frac{\omega}{v_\perp}\right)^2}.$$

Power spectra derived from this theory have been stated elsewhere: Ishimaru (1978, Chapt. 19, Eq. 16) uses the Kolmogorov spectrum

$$\Phi_n(\kappa) = 0.033 C_n^2 \kappa^{-11/3}$$

to derive the power spectrum of phase fluctuations, where C_n^2 is the phase *structure constant* that determines the intensity of the fluctuations. Tatarskii (1971, §52, Eq. 34) used a modified Kolmogorov spectrum

$$\Phi_n(\kappa) = 0.033 C_n^2 \kappa^{-11/3} \exp(-\kappa^2/\kappa_m^2), \quad (\text{G.4})$$

to also derive phase fluctuations. Tatarski's modified spectrum includes an exponential taper which represents turbulence with a finite *inner* scale length determined by κ_m . His results are also presented by Lawrence and Strohbehn (1970, Eq. 44). That equation may be evaluated at asymptotic limits with reference to

$$\Omega = \frac{f}{v_\perp} (2\pi\lambda L)^{1/2},$$

and the result is

$$W_s(f) = 3.28 \cdot 10^{-2} C_n^2 k^2 L v_\perp^{5/3} f^{-8/3}, \quad \Omega \ll 1, \quad (\text{G.5})$$

$$W_s(f) = 1.64 \cdot 10^{-2} C_n^2 k^2 L v_\perp^{5/3} f^{-8/3}, \quad \Omega \gg 1. \quad (\text{G.6})$$

G.3.2 Temporal power spectra in a smoothly varying medium

For astronomical applications it is unreasonable to assume that the medium is homogeneous, since the pathlength of propagation will be many times larger than the outer scale size of turbulent eddies. Tatarski (1961, Chapt. 8) and Tatarskii (1971, §48) uses the following substitution for waves that propagate through slowly varying media. The power spectrum of index of refraction fluctuations is now written

$$\Phi_n(h, \kappa) = C_n^2(h) \Phi_n^0(\kappa).$$

The spectral density of $\Phi_n^0(\kappa)$ remains unchanged along the path of propagation h , but the total power contained in the spectrum varies according to the structure function. This substitution is also used by Lawrence and Strohbehn in their review (1970). The covariances of log-amplitude and phase therefore can be written:

$$B_\chi(L, \rho) = 4\pi^2 k^2 \int_0^L C_n^2(h) dh \int_0^\infty \kappa d\kappa J_0(\kappa\rho) \sin^2 \left[\frac{L-h}{2k} \kappa^2 \right] \Phi_n^0(\kappa),$$

$$B_s(L, \rho) = 4\pi^2 k^2 \int_0^L C_n^2(h) dh \int_0^\infty \kappa d\kappa J_0(\kappa\rho) \cos^2 \left[\frac{L-h}{2k} \kappa^2 \right] \Phi_n^0(\kappa).$$

The next simplifying assumption is to use the geometrical optics approximation, $\kappa^2 L/k \ll 1$, so that the sine squared term becomes an angle squared and the cosine term approaches unity.

$$B_\chi(L, \rho) = 4\pi^2 k^2 \int_0^L C_n^2(h) dh \int_0^\infty \kappa d\kappa J_0(\kappa\rho) \frac{(L-h)^2}{4k^2} \kappa^4 \Phi_n^0(\kappa),$$

$$B_s(L, \rho) = 4\pi^2 k^2 \int_0^L C_n^2(h) dh \int_0^\infty \kappa d\kappa J_0(\kappa \rho) \Phi_n^0(\kappa).$$

This same approach can be used to write the temporal power spectra.

G.4 Temporal Power Spectra of Everything

If the geometric optics approximation is used, where $f_s(\kappa) \simeq 2$, then the equation for the power spectrum of phase fluctuations is greatly simplified. For fluctuations in a *locally* homogeneous medium we have, from Eq. G.3,

$$W_s(\omega) = \frac{16\pi^2 k^2 L}{v_\perp} \int_0^\infty \Phi_n(\kappa) d\kappa'.$$

The total power spectrum as seen from the ground can then be derived by integrating $W_s(\omega)$ as a function of height through the atmosphere, taking into account changes of the wind speed and phase structure constant. Colavita uses this framework to describe various power spectra relevant to astrometric observations (Colavita, 1985; Colavita, 1987). He uses the correlation function of phase to define other correlation functions, including phase-difference, relative phase-difference, and time-lagged relative phase-difference, for instances where the wind velocity is parallel or perpendicular to the interferometer's baseline. These new correlation functions modify the power spectrum according to basic Fourier transform relationships, and appear as a multiplicative factor, $g(\kappa, \omega)$, in the equation.

$$W_s(\omega) = 16\pi^2 k^2 \int_0^\infty dh \frac{1}{v_\perp} \int_0^\infty d\kappa' \Phi_n(h, \kappa) g(\kappa', \omega).$$

They can also be thought of as filter functions, in the same sense as was used in Eq.G.2.

G.5 Summary

The theory of wave propagation through the atmosphere has been described using the Rytov approximation and the spectral representation of random variables. The models of index of refraction fluctuations, $\Phi_n(\kappa)$, have not been described in detail, but the formulations have been presented so that they may be applied to the study of the power spectra of phase fluctuations. The theory is described further in Chapt. 8 where phase-difference power spectra are described.

Bibliography

Adaptive Optics Associates, Inc., (1988) *AOA Photon Digitizing Camera* May 21, 1988 (Adaptive Optics Associates Inc: 54 Cambridge Park Dr., Cambridge, MA 02140-2308) A User's Manual.

R.W. Airy, T.J. Norton, B.L. Morgan, J.L.A. Fordham, D.A. Bone, and J.R. Powell, (1990) "MCP image intensifier with improved DQE," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **1243**, 140–148.

J.T. Armstrong et al., (1992) "The orbit of α Equulei measured with long-baseline optical interferometry: component masses, spectral types, and evolutionary state," *Astron. J.*, **104**, 241–252.

Austek Microsystems, (1988a) *A User Guide for the A41102* (Austek Microsystems Pty. Ltd.: Adelaide) .

Austek Microsystems, (1988b) *A41102 Product Specification* (Austek Microsystems Pty. Ltd.: Adelaide) .

G.R. Ayers, M.J. Northcott, and J.C. Dainty, (1988) "Knox-Thompson and triple-correlation imaging through atmospheric turbulence," *J. Opt. Soc. Am. A*, **5**, 963–986.

J.W. Beletic, R.M. Goody, and D.J. Tholen, (1989) "Orbital Elements of Charon from Speckle Interferometry," *Icarus*, **79**, 38–46.

J.W. Beletic and R.M. Goody, (1992) "Recovery of planetary images by speckle imaging," *Applied Optics*, **31**, 6909–6921.

G.D. Bergland, (1969) "A guided tour of the fast Fourier transform," *IEEE Spectrum*, **6**, 41–52.

M. Bester, W.C. Danchi, C.G. Degiacomi, L.J. Greenhill, and C.H. Townes, (1992) "Atmospheric fluctuations: empirical structure functions and projected performance of future instruments," *Ap. J.*, **392**, 357–374.

A.J. Booth, Y.A. Gilliland, J. Davis, W.J. Tango, S.M. Owens and N. Brown, (1992) 'SUSI's Path Compensator System' in *High Resolution Imaging by Interferometry II*. (European Southern Observatory: Garching bei München) 1127–1132.

M. Born & E. Wolf, (1980) *Principles of Optics* 6th Edition (Pergamon Press: Oxford) §10.3.1.

S.M. Bozic, (1979) *Digital and Kalman Filtering* Part 2–Optimum linear estimation (Edward Arnold: London) 81ff.

R.N. Bracewell, (1986) *The Fourier Transform and its Applications* (McGraw-Hill Book Company: Singapore) International Edition.

T. ten Brummelaar, (1992) ‘An active wavefront tilt correction servo for SUSI’ in *High Resolution Imaging by Interferometry II*. (European Southern Observatory: Garching bei München) 1075–1078.

A.E. Bryson & Y.C. Ho, (1969) *Applied Optimal Control* Chapt. 12 (Ginn: Waltham MA) 348–389.

D. Buscher, (1988) *Getting the most out of COAST* (Cambridge University: November 1988) Ph.D. thesis.

D. Buscher, (1989) ‘Low light level limits to tracking atmospheric fringe wander.’ in *Quantum limited imaging and information processing*. 1989 Technical Digest Series Vol. 13, Conf. Edition. (OSA and Air Force Office of Scientific Research: North Falmouth MA) .

D.F. Buscher, J.T. Armstrong, D. Mozurkewich, C.D. Denison, M.M. Colavita and M. Shao, (1992) ‘Atmospheric fluctuation measurements with the MkIII interferometer and the power spectra of everything’ in *High Resolution Imaging by Interferometry II*. (European Southern Observatory: Garching bei München) 1029–1038.

A. Cañas, (1989) *A prototype design for the analogue electronics of a PAPA camera* (Applied Optics Department, Blackett Laboratories: Imperial College of London) An unpublished report containing suggestions for the design of the new analog board.

B.G. Clark, (1989) ‘Coherence in Radio Astronomy’ in *Synthesis Imaging in Astronomy* Edited by R.A. Perley, F.R. Schwab, and A.H. Bridle (Brigham Young University: Provo, Utah) .

L.D. Clark, M. Shao and M. Colavita, (1986) “A Photon-Camera Star Tracker for Stellar Interferometry,” *Proc. Soc. Photo-Opt. Instrum. Eng.*, **627**, 838–845.

S.F. Clifford, (1978) ‘The classical theory of wave propagation in a turbulent medium’ in *Laser Beam Propagation in the Atmosphere* (Springer-Verlag: Berlin) 9–43.

W.T. Cochran et al., (1967) “What is the Fast Fourier Transform?,” *IEEE Trans. Audio Electroacoust.*, **15**, 45–55.

M.M. Colavita, (1985) *Atmospheric Limitations of a Two-Color Astrometric Interferometer* (Thesis: Massachusetts Institute of Technology) 75–102.

M.M. Colavita, M. Shao, and D.H. Staelin, (1987) “Atmospheric phase measurements with the Mark III stellar interferometer,” *Appl. Opt.*, **26**, 4106–4112.

M.M. Colavita, M. Shao, and D.H. Staelin, (1987) “Two-color method for optical astrometry: theory and preliminary measurements with the Mark III stellar interferometer,” *Appl. Opt.*, **26**, 4113–4122.

M.M. Colavita and M. Shao, (1988) Photon-starved astrometric measurements with a large-aperture interferometer in *ESO Conference Proceedings on ‘High Resolution Imaging by Interferometry’* 833–839 (European Southern Observatory: Garching bei München) Fritz Merkle, Editor.

P. Connes, (1970) “Astronomical Fourier Spectroscopy,” *Ann. Rev. Astron. Astrop.*, **8**, 209–230.

M. B. Corbett, (1992) “Microchannel Plates: Principles of Operation,” *The Photonics Handbook*, **3**, H-102–105.

C.E. Coulman, (1985) “Fundamental and Applied Aspects of Astronomical Seeing,” *Ann. Rev. Astron. Astrophys.*, **23**, 19–57.

- G.C. Cox, (1992) 'COAST—Acquisition and auto-guider system' in *High Resolution Imaging by Interferometry II*. (European Southern Observatory: Garching bei München) 929–933.
- K. Creath, (1988) "Phase-measurement interferometry techniques," *Progress in Optics*, **26**, 349–393.
- I.P. Csorba, (1985) *Image Tubes* (Howard W. Sams & Co. Inc.: Indianapolis) .
- J.C. Dainty, (1971) "Detection of images immersed in speckle noise," *Optica Acta*, **18**, 327–339.
- J.C. Dainty and A.H. Greenaway, (1979) "Estimation of spatial power spectra in speckle interferometry," *J. Opt. Soc. Am.*, **69**, 786–790.
- J.C. Dainty, (1984) *Laser Speckle and Related Phenomena* (Springer-Verlag: Berlin) 255–320.
- W.B. Davenport, Jr., and W.L. Root, (1958) *An Introduction to the Theory of Random Signals and Noise* (McGraw-Hill: New York) 54–55.
- J. Davis, (1985a) 'Measuring stars with high angular resolution: current status and future prospects' in *Calibration of Fundamental Stellar Quantities* (IAU Symposium: D.S. Hayes et al. editors) 193–208.
- J. Davis, (1985b) "A new very high angular resolution stellar interferometer," *Proc. Astr. Soc. Aust.*, **6**, 38–43.
- J. Davis & W.J. Tango, (1985) "The Sydney University 11.4 m Prototype Stellar Interferometer," *Proc. Astr. Soc. Aust.*, **6**, 34–38.
- J. Davis & W.J. Tango, (1986) "New determination of the angular diameter of Sirius," *Nature*, **323**, 234–235.
- J. Davis, W.J. Tango, A.J. Booth, R.A. Minard, T. ten Brummelaar, and R.R. Shobbrook, (1992) 'An Update on SUSI' in *High Resolution Imaging by Interferometry II*. (European Southern Observatory: Garching bei München) 741–745.
- L.M. Smith & C.C. Dobson, (1989) "Absolute displacement measurements using modulation of the spectrum of white light in a Michelson interferometer," *Appl. Opt.*, **28**, 3339–3342.
- W.G. Driscoll, Ed., (1978) *Handbook of Optics* (McGraw-Hill Book Company: New York) 2–56, 2–57.
- J. Drummond, A. Eckart, and E.K. Hege, (1988) "Speckle Interferometry of Asteroids: IV. Reconstructed Images of 4 Vesta," *Icarus*, **73**, 1–14.
- S.M. Ebstein, (1987) *Speckle Imaging of Active Galactic Nuclei: NGC 1068 and NGC 4151* (Ph.D. Thesis: Harvard University) 6–14.
- S.M. Ebstein, N.P. Carleton, and C. Papaliolios, (1989) "Speckle Imaging of NGC 1068 and NGC 4151 in the [O III] 5007 Å Line and Nearby Continuum," *Astrophys. J.*, **336**, 103–111.
- R.L. Fante, (1975) "Electromagnetic Beam Propagation in Turbulent Media," *IEEE Trans. Antennas Propagat.*, **63**, 1669–1692.
- R. Foy, (1988) *Instrumentation for Ground-Based Optical Astronomy* 588–592 (Springer-Verlag: New York) The Photon Counting Camera CP40.
- D.L. Fried, (1965) "Statistics of a geometric representation of wavefront distortion," *J. Opt. Soc. Am.*, **55**, 1427–1435.

- M. Gardner, (1972) "The Curious Properties of the Gray Code," *Sci. Am.*, **227**, No. 2, 106–109.
- T. Gonsiorowski, (1984) "Variable Threshold Discrimination in a Photon-Imaging Detector," *Appl. Opt.*, **23**, 1060–1065.
- T. Gonsiorowski, (1986) "A New Product for Photon-Limited Imaging," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **627**, 626–630.
- J.W. Goodman and J.F. Belsher, (1976a) 'Photon limited images and their restoration,' RADC-TR-76-50 (Rome Air Development Center: Griffiss AFB, NY, 13441) .
- J.W. Goodman and J.F. Belsher, (1976b) 'Precompensation and postcompensation of photon limited degraded images,' RADC-TR-76-382 (Rome Air Development Center: Griffiss AFB, NY, 13441) .
- J.W. Goodman and J.F. Belsher, (1977) 'Photon limitations in imaging and restoration,' RADC-TR-77-175 (Rome Air Development Center: Griffiss AFB, NY, 13441) .
- J.W. Goodman, (1984) *Statistical Optics* (McGraw Hill Book Company: New York) 490–501.
- D.P. Greenwood and D.L. Fried, (1976) "Power spectra requirements for wave-front-compensative systems," *J. Opt. Soc. Am.*, **66** , 193–206.
- D.P. Greenwood, (1977) "Bandwidth specification for adaptive optics systems," *J. Opt. Soc. Am.*, **67** , 390–393.
- F.J. Harris, (1978) "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, **66**, 51–83.
- F.G. Heath, (1972) "Origins of the Binary Code," *Sci. Am.*, **227**, No. 2, 76–83.
- E. Hecht, (1987) *Optics* (Addison Wesley: Reading MA) Chapt. 12.
- C.B. Hogge and R.R. Butts, (1976) "Frequency spectra for the geometrical representation of wavefront distortions due to atmospheric turbulence," *IEEE Trans. Antennas Propagat.*, **24**, 144–154.
- Interrupt Systems, (1988) *Avenue Electronics 68000 Single Board Computer Model No. AV68K User Manual* (9 Robinlee Ave: Burwood East 3151, Australia) .
- A. Ishimaru, (1978) *Wave Propagation and Scattering in Random Media* (Academic Press: New York) 346–375, 388–406.
- A.J. Jerri, (1977) "The Shannon sampling theorem—its various extensions and applications: a tutorial review," *Proc. IEEE*, **65**, 1565–1596.
- J.H. Justice, N.L. Owsley, J.L. Yen, and A.C. Kak, (1985) *Array Signal Processing* Chapt. 5 (Prentice-Hall: Englewood Cliffs, NJ) .
- R.E. Kalman & R.S. Bucy, (1961) "New results in linear filtering and prediction theory," *ASME, J. Basic Eng.*, **83D**, 95–108.
- M. Karovska, P. Nisenson, and R. Noyes, (1986) "On the Alpha Orionis Triple System," *Astrophys. J.*, **308**, 260–269.
- M. Karovska, P. Nisenson, and R.V. Stachnik, (1986) "Resolution of the Halo Binary μ Cas at Optical Wavelengths," *Astronom. J.*, **92**, 898–902.
- M. Karovska, L. Koechlin, P. Nisenson, C. Papaliolios, and C. Standley, (1989) "Measurements of the Diameter of the Large Magellanic Cloud Supernova SN 1987A," *Astrophys. J.*, **340**, 435–442.

- S.M. Kay & S.L. Marple, Jr., (1981) "Spectrum analysis—a modern perspective," *Proc. IEEE*, **69**, 1380–1418.
- E.J. Kim, (1989) *Dispersed fringe group delay astrometry using the Mark III stellar interferometer* (Mass. Inst. of Technology: Cambridge MA) M.S.E.E. thesis.
- L. Koechlin, (1985) "Interférométrie stellaire dans l'espace: détection des franges," *J. Optics (Paris)*, **16**, 269–276.
- S.R. Kulkarni, S. Prasad, and T. Nakajima, (1991) "Noise in optical synthesis images. II. Sensitivity of an nC_2 interferometer with bispectrum imaging," *J. Opt. Soc. Am. A*, **8**, 499–510.
- A. Labeyrie, (1970) "Attainment of diffraction limited resolution in large telescopes by Fourier analysing speckle patterns in star images," *Astron. Astrophys.*, **6**, 85–87.
- A. Labeyrie, (1978) "Stellar Interferometry Methods," *Ann. Rev. Astron. Astrophys.*, **16**, 77–102.
- M. Lampton, (1981) "The Microchannel Image Intensifier," *Sci. Am.*, **245**, No. 5, 46–55.
- B. Laprade and M. Wheeler, (1990) "Dynamic Range Optimization in Microchannel Plates," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **1243**, 132–139.
- D.W. Latham, (1982) 'Spectroscopy with photon counting Reticons' in *Instrumentation with Large Optical Telescopes*, Proc. IAU No. 67, C.M. Humphries, Ed. (Cambridge University Press: New York) 259–270.
- R.S. Lawrence and J.W. Strohbehn, (1970) "A Survey of Clear-Air Propagation Effects Relevant to Optical Communications," *IEEE Trans. Antennas Propagat.*, **58**, 1523–1545.
- P.R. Lawson, (1993) *Hardware documentation and User's Manual for the GDT processors* (Chatterton Astronomy Department: Sydney University) .
- R.W. Lee and J.C. Harp, (1969) "Weak Scattering in Random Media, with Applications to Remote Probing," *Proc. IEEE*, **57**, 375–406.
- B. Leskovar, (1977) "Microchannel Plates," *Phys. Today*, **30**, No. 11, 42–49.
- P.A. Mandics, R.W. Lee, and A.T. Waterman, Jr., (1973) "Spectra of short-term fluctuations of line-of-sight signals: Electromagnetic and acoustic," *Radio Science*, **8**, 185–201.
- H.A. McAlister, (1985) "High angular resolution measurements of stellar properties," *Ann. Rev. Astron. Astrophys.*, **23**, 95–87.
- A.A. Michelson & F.G. Pease, (1921) "Measurement of the diameter of α Orionis with the interferometer," *Ap. J.*, **53**, 249–259.
- J.S. Morgan, (1988) 'Speckle imaging with the MAMA detector' in *High Resolution Imaging by Interferometry*. (European Southern Observatory: Garching bei München) 381–391.
- D. Mozurkewich et al., (1991) "Angular diameter measurements of stars," *Astron. J.*, **101**, 2207–2219.
- D. Mozurkewich, (1993) Personal communication (IAU Conf. 158: Sydney, Australia) .
- National Instruments, (1990) *AT-DIO-32F User Manual* (National Instruments Corporation: Austin TX 1990) Part Number 320147-01.
- N.S. Nightingale and D.F. Buscher, (1991) "Interferometric seeing measurements at the La Palma Observatory," *Mon. Not. R. Astr. Soc.*, **251**, 155–166.

- P. Nisenson & W. Traub, (1987) 'Magnitude limit of the group delay tracking method for long baseline interferometry' in *Interferometric Imaging in Astronomy* (National Optical Astronomy Observatories: Tucson AZ) 129–133.
- P. Nisenson, R.V. Stachnik, M. Karovska, and R. Noyes, (1985) "A New Optical Source Associated with T Tauri," *Astrophys. J.*, **297**, L17–20.
- P. Nisenson, C. Papaliolios, M. Karovska, and R. Noyes, (1987) "Detection of a Very Bright Source Close to the LMC Supernova SN 1987A," *Astrophys. J.*, **320**, L15–L18.
- T. Nolan, (1990) "Real-time data acquisition using DMA," *Dr. Dobb's Journal*, **15**, No. 1, 28–37, 94–96.
- R.J. Noll, (1976) "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.*, **66**, 207–211.
- C. Papaliolios and L. Mertz, (1982) "New Two-Dimensional Photon Camera," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **331**, 360–364.
- C. Papaliolios, P. Nisenson, and S. Ebstein, (1985) "Speckle Imaging with the PAPA Camera," *Appl. Opt.*, **24**, 287–292.
- C. Papaliolios, M. Karovska, L. Koechlin, P. Nisenson, C. Standley, and S. Heathcote, (1989) "Asymmetry of the Envelope of Supernova 1987A," *Nature*, **338**, 565–566.
- R.A. Perley, F.R. Schwab, and A.H. Bridle, (1989) *Synthesis Imaging in Radio Astronomy* (Brigham Young University: Provo, Utah) 117–138.
- Plessey Semiconductors, (1988) *Digital Signal Processing IC Handbook* (Pub. No. PS2252: Swindon, UK) .
- H.K. Pollehn, (1980) *Applied Optics and Optical Engineering, Vol. VI* "Image Intensifiers", 393–437 (Academic Press: New York) Chapt. 10.
- S. Prasad and S.R. Kulkarni, (1989) "Noise in optical synthesis images. I. Ideal Michelson interferometer," *J. Opt. Soc. Am. A*, **6**, 1702–1714.
- W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, (1992) *Numerical Recipes in C* Second edition (Cambridge University Press: New York) 894–896.
- W.H. Press & G.B. Rybicki, (1989) "Fast algorithm for spectral analysis of unevenly sampled data," *Ap. J.*, **338**, 277–280.
- J.G. Proakis and D.G. Manolakis, (1992) *Digital Signal Processing* (Maxwell MacMillan: New York) .
- Y. Rabbia, (1988) 'Attempts for optical path difference measurements at I2T/CERGA' in *High Resolution Imaging by Interferometry*. (European Southern Observatory: Garching bei München) 705–714.
- L.R. Rabiner & C.M. Rader, Editors, (1972) *Digital Signal Processing* (IEEE Inc: New York) A volume in the IEEE Press Selected Reprint Series.
- C.M. Rader and B. Gold, (1967) "Digital Filter Design Techniques in the Frequency Domain," *Proc. IEEE*, **55**, 149–171.
- R.D. Reasenberg, (1990) "Kalman-filter fringe tracking in an optical interferometer," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **1237**, 172–182.
- S.O. Rice, (1944) "Mathematical Analysis of Random Noise," *Bell System's Technical Journal*, **23**, 282–332.

- S.O. Rice, (1945) "Mathematical Analysis of Random Noise, Part II," *Bell System's Technical Journal*, **24**, 46–156.
- S.O. Rice, (1948) "Statistical Properties of a Sine Wave Plus Random Noise," *Bell System's Technical Journal*, **Jan.**, 109–157.
- F. Roddier, (1981) "The Effects of Atmospheric Turbulence in Optical Astronomy," *Progress in Optics*, **19**, 281–376.
- F. Roddier, (1986) "Pupil plane versus image plane in Michelson stellar interferometry," *J. Opt. Soc. Am. A*, **3**, 2160–2166.
- G. Schumacher, P. Cruzalèbes, A. Jeanne, and Y. Rabbia, (1992) 'Status of the CHARON optical interferometer project' in *High Resolution Imaging by Interferometry II* (European Southern Observatory: Garching bei München) 715–719.
- M. Shao & D.H. Staelin, (1977) "Long-baseline optical interferometer for astrometry," *J. Opt. Soc. Am.*, **67**, 81–86.
- M. Shao & D.H. Staelin, (1980) "First fringe measurements with a phase tracking stellar interferometer," *Appl. Opt.*, **19**, 1519–1522.
- M. Shao et al., (1988) "The Mark III stellar interferometer," *Astron. Astrophys.*, **193**, 357–371.
- M. Shao & M. Colavita, (1992) "Long-baseline optical and infrared stellar interferometry," *Ann. Rev. Astron. Astrophys.*, **30**, 457–498.
- D. C. Slater, J. G. Timothy, J. S. Morgan, and D. B. Kasse, (1990) "Imaging MAMA Detector Systems," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **1243**, 35–49.
- W.J. Smith, (1990) *Modern Optical Engineering: The Design of Optical Systems* (McGraw-Hill Book Company: New York) 96–99.
- H.W. Sorenson, (1970) "Least-squares estimation: from Gauss to Kalman," *IEEE Spectrum*, **7**, 63–68.
- H.W. Sorenson, (1983) "Editorial: On Applications of Kalman Filtering," *IEEE Trans. Automat. Contr.*, **28**, 254–255.
- H.W. Sorenson, Editor, (1985) *Kalman Filtering: Theory and Applications* (IEEE Press: New York) A volume in the Selected Reprint Series.
- B. Sorrente, F. Cassaing, D. Beal, E. Prieto, Y. Rabbia, M. Dugue, (1992) 'Active phase stabilization with ASSI: Results and perspectives' in *High Resolution Imaging by Interferometry II*. (European Southern Observatory: Garching bei München) 1133–1142.
- C. Standley and P. Nisenson, (1989) "A Second Generation PAPA Camera," *Bull. Am. Astron. Soc.*, **21**, 1072.
- W.H. Steele, (1987) *Interferometry* (Cambridge University Press: Cambridge) .
- J.W. Strohbehn, Editor, (1978) *Laser Beam Propagation in the Atmosphere* (Springer-Verlag: Berlin) 1–106.
- W.J. Tango, (1979) 'The Monteporzio two meter amplitude interferometer' in Proc. IAU Colloq. No. 50, *High Angular Resolution Stellar Interferometry*. (University of Sydney: Sydney Australia) .
- W.J. Tango & R.Q. Twiss, (1980) "Michelson stellar interferometry," *Progress in Optics*, **17**, 239–277.

- W.J. Tango, (1990) "Dispersion in stellar interferometry," *Appl. Opt.*, **29**, 516–521.
- V.I. Tatarski, (1961) *Wave Propagation in a Turbulent Medium* (McGraw-Hill Book Company: New York) 128–150.
- V.I. Tatarskii, (1971) *The Effects of the Turbulent Atmosphere on Wave Propagation* (Israel Program for Scientific Translations: Jerusalem) §52.
- V.I. Tatarskii and V.U. Zavorotnyi, (1980) "Strong fluctuations in light propagation in a randomly inhomogeneous medium," *Progress in Optics*, **18**, 204–256.
- A.R. Thompson and R.N. Bracewell, (1974) "Interpolation and Fourier transformation of fringe visibilities," *Astron. J.*, **79**, 11–24.
- A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., (1986) *Interferometry and Synthesis in Radio Astronomy* (John Wiley & Sons: New York) .
- J.G. Timothy, (1985) "Multinode Microchannel Array Detector Systems: Performance Characteristics," *Optical Engineering*, **24**, 287.
- W.A. Traub, M.G. Lacasse, and N.P. Carlton, (1990a) "Spectral dispersion and fringe detection in IOTA," *Proc. Soc. Photo-Opt. Instrum. Eng.*, **1237**, 145–152.
- W.A. Traub, (1990b) "Constant-dispersion grism spectrometer for channeled spectra," *J. Opt. Soc. Am.*, **7**, 1779–1791.
- F. Vakili, Y. Rabbia and I. Percheron, (1988) 'Image reconstruction in optical long baseline interferometry using spectral visibility fitting' in *High Resolution Imaging by Interferometry*. (European Southern Observatory: Garching bei München) 769–779.
- J.F. Walkup and J.W. Goodman, (1973) "Limitations of fringe parameter estimation at low light levels," *J. Opt. Soc. Am.*, **63**, 399.
- Nelson Wax, Editor, (1954) *Selected Papers on Noise and Stochastic Processes* (Dover Publications: New York) 133–294.
- P.D. Welch, (1969) "A Fixed-Point Fast Fourier Transform Error Analysis," *IEEE Trans. Audio Electroacoust.*, **17**, 151–157.
- Anthony D. Whalen, (1971) *Detection of Signals in Noise* (Academic Press: New York) 99–109.
- J. L. Wiza, (1979) "Microchannel Plate Detectors," *Nucl. Instrum. Methods*, **162**, 587–601.
- J. Wyant, (1975) "Use of an ac heterodyne lateral shear interferometer with real-time wavefront correction systems," *Appl. Opt.*, **14**, 2622.
- J.L. Yen, (1956) "On nonuniform sampling of bandwidth-limited signals," *IRE Trans. Circuit Theory*, **Dec.**, 251–257.